DUKE MATHEMATICAL JOURNAL

Volume 9, Number 1

MARCH, 1942

DUKE UNIVERSITY PRESS

DURHAM, N. C.

DUKE MATHEMATICAL JOURNAL

This periodical is published quarterly under the auspices of Duke University by Duke University Press at Durham, North Carolina. It is printed at Mt. Royal and Guilford Avenues, Baltimore, Maryland, by the Waverly Press.

Entered as second class matter at the Post Office, Durham, North Carolina. Additional entry at the Post Office, Baltimore, Maryland.

The subscription price for the current year is four dollars, postpaid; back volumes, five dollars each, carriage extra. Subscriptions, orders for back numbers, and notice of change of address should be sent to Duke University Press, Durham, North Carolina.

Individual and institutional members of the Mathematical Association of America may subscribe to the current volume at half price. To get the reduced price, orders for subscriptions must bear the mention "Member MAA." If an order at the reduced price is placed through an agent, the purchaser must pay any commission charge incurred.

Since 1935 the Mathematical Association of America has given the Duke Mathematical Journal an annual subsidy, in return for which the half-rate has been allowed. Having served its purpose of aiding the establishment of the Journal, the subsidy is to be discontinued at the end of 1942. In view of the help already received from the Association, however, Duke University Press will for at least five years continue to allow the half-rate to any one who in 1942 is a subscriber at the reduced rate provided his subscription to the Journal and membership in the Association remain unbroken. This arrangement is expected to be permanent, but the Press reserves the right to modify or withdraw it after the five years and to change the basic rate at the beginning of any calendar year.

Manuscripts and editorial correspondence should be addressed to Duke Mathematical Journal, 4785 Duke Station, Durham, North Carolina. An Author's Manual containing detailed information about the preparation of papers for publication will be sent on request.

Authors are entitled to one hundred free reprints. Additional copies will be supplied at cost. All reprints will be furnished with covers unless the contrary is specifically requested.

The American Mathematical Society is officially represented on the Editorial Board by Professors Murray and Ward.

Made in United States of America

WAVERLY PRESS, INC.
BALTIMORE, U. S. A.

A COMPARISON OF LINEAR MEASURES IN THE PLANE

By Seymour Sherman

In generalizing from the notion of the length of curve and linear measure of a linear set to the linear measure of a plane set the following considerations arise:

1. Does the new measure give the expected results for point sets which can be treated by the old methods?

2. Is the new measure invariant under Euclidean transformation of the set?

3. Does the new measure have the usual measure properties, i.e., is it completely additive; does it satisfy the general Carathéodory measure postulates?

Some such generalizations satisfying these and more subtle[1] requirements have been proposed by Carathéodory, Gross, Steinhaus,[2] Favard, Kolmogoroff,[3] Appert, Randolph, and Morse. Of these measures the ones associated with Carathéodory, Gross, Appert, Randolph, and Morse are closely related and involve countable decompositions of the given set while the ones suggested by Kolmogoroff and Steinhaus diverged into different paths. Kolmogoroff measure originated, in part, with the notion of Schmidt [10] that the measure of a contracted set is not greater than the measure of the set. Steinhaus measure originated (1) with the measure (see footnote 6) of sets of lines used for the Buffon needle problem and (2), surprisingly enough, with a mechanical device[4] for measuring lengths of curves as seen under a microscope.

We are mainly concerned with the relationship between Carathéodory linear measure and Steinhaus linear measure. We complete Steinhaus' proof that the Steinhaus measure of a rectifiable Jordan curve is equal to twice its length as defined by the inscribed polygon approach. In Theorem 4, we show that, contrary to Steinhaus' expressed belief,[5] there are sets (irregular in the sense of Besicovitch) whose Steinhaus measure is different from twice their Carathéodory linear measure, and, in general, if a set is measurable Carathéodory, then its Steinhaus linear measure is equal to twice the Carathéodory linear measure of its regular part. Unlike other linear measures, Steinhaus linear measure may

Received January 2, 1941. The author is indebted to Prof. J. F. Randolph for many helpful suggestions.

[1] One such requirement is that the outer linear Lebesgue measure of the projection of the set on any line be not greater than the outer measure of the set. For a discussion of this requirement see [5]. Numbers in square brackets refer to the bibliography. For considerations involving a natural generalization to higher dimensions see [7] and [8].

[2] See [12] and [13]. This measure is completely additive over the family of sets measurable Steinhaus but, since it is not introduced by means of an exterior measure function, the Carathéodory measure postulates do not apply. This measure was later independently suggested by Favard [4].

[3] Kolmogoroff measure [6] is defined merely for analytic sets and so the general Carathéodory postulates do not apply. It is completely additive over analytic sets.

[4] See [12].

[5] See [13], p. 354.

be applied not only to find a linear measure for a point set but also to find a length for a parametrized curve. In Theorem 5 we establish by means of Steinhaus measure a very natural relation between the length of a rectifiable curve and the Carathéodory measures of the sets of its multiple points.

Deltheil measure.[6] Let line l in the plane have coordinates (ρ, θ) assigned to it in the following manner. If l is not vertical and does not pass through the origin or if l is vertical and lies to the left of the origin, if ρ $(\rho > 0)$ is the distance from the origin to the line and θ $(2\pi > \theta > 0)$ is the angle made between the horizontal ray through the origin and the normal from the origin to l, then (ρ, θ) are the coordinates of l; if l is vertical and passes through the origin, then l has the coordinates $(0, 0)$, $(0, \pi)$, and $(0, 2\pi)$; if l is vertical and lies to the right of the origin, and ρ $(\rho > 0)$ is the distance from the origin to l, then l has the coordinates $(\rho, 0)$ and $(\rho, 2\pi)$; and if l is not vertical and does pass through the origin and θ $(\pi > \theta > 0)$ is the angle made between the initial line and the undirected normal, then l has the coordinates $(0, \theta)$ and $(0, \theta + \pi)$. If S is a set of lines such that

$$D(S) = \iint_{(\rho,\theta) \in S} d\rho \, d\theta, \tag{1}$$

where the integration is in the sense of Lebesgue, exists, then we say that S is measurable (D) and $D(S)$ is its Deltheil measure. This measure is invariant with respect to Euclidean transformations of S.

Steinhaus measure. If A is a set of lines in the plane, we define $f_A(\rho, \theta)$ to be a function of lines whose value is the number of points of A on the line (ρ, θ). Thus $f_A(\rho, \theta)$ has the value 0, n (positive integer), or $+\infty$. If $f_A(\rho, \theta)$ is integrable Lebesgue, let

$$St(A) = \iint f_A(\rho, \theta) \, d\rho \, d\theta, \tag{2}$$

where the integration is taken over the whole (ρ, θ)-plane. This Steinhaus measure is invariant under Euclidean transformations of A. It is easy to see that the Steinhaus measure of a line segment is equal to twice its length and that the Steinhaus measure of a polygon is equal to twice its perimeter. Steinhaus stated that if A is a continuous rectifiable curve, then

$$St(A) = 2l(A), \tag{3}$$

where $l(A)$ is the length of A calculated by polygonal approximation and $St(A)$ is calculated by giving to each point of A the multiplicity it gains by the parametrization. After a few definitions and lemmas we shall give the complete proof.

[6] See R. Deltheil, *Probabilités Géométriques*, Paris, 1926.

For a plane set A,

$$T(A) = \underset{df(\rho,\theta)}{E}\,[f_A(\rho, \theta) > 0], \tag{4}$$

$$C_{T(A)}(\rho, \theta) = {}_{df}\text{characteristic function of } T(A), \tag{5}$$

and if (R, α) are the polar coordinates of a point in A, then

$$T(R, \alpha) = \underset{(\rho,\theta)}{E}\,[\rho = R\cos(\alpha - \theta), \rho \geqq 0, 2\pi \geqq \theta \geqq 0]. \tag{6}$$

It is easy to see that

$$T(A) = \sum_{(R,\alpha)\,\epsilon\,A} T(R, \alpha) \tag{7}$$

and

$$\sum_{j\,\epsilon\,E} T(A^j) = T\sum_{j\,\epsilon\,E} A^j, \tag{8}$$

where the A^j are plane sets and the index j can have any range E whatsoever.

LEMMA 1.[7] *If $m_C(A) = 0$, then $St(A) = 0$.*

Proof. If $m_C(A) = 0$, then there exists a sequence $\{\mathfrak{C}_i\}$ of countable coverings $\{\mathfrak{C}_i\}$ of A by squares such that $S_i \to 0$ as $i \to \infty$ where S_i is the sum of the sides of the squares in the covering $\mathfrak{C}_i$. Obviously

$$T(A) \subset T(\mathfrak{C}_i),$$

where $T(\mathfrak{C}_i)$ is the union of the transforms of the squares in the covering $\mathfrak{C}_i$, and[8]

$$|\,T(A)\,|_2 = |\,T(\mathfrak{C}_i)\,|_2 \to 0.$$

Thus $T(A)$ is measurable (Lebesgue), $|\,T(A)\,|_2 = 0$, and $\iint f_A(\rho, \theta)\,d\rho\,d\theta = 0$.

LEMMA 2. *If C is a continuous rectifiable curve parametrized according to arc length, then*

$$\left|\,\underset{s}{E}\left[\frac{dx}{ds}\text{ fails to exist or }\frac{dy}{ds}\text{ fails to exist or both }\frac{dx}{ds}\text{ and }\frac{dy}{ds}\text{ equal zero}\right]\right|_1 = 0,$$

and the Deltheil measure of those lines which are tangent to the curve at points at which $\frac{dx}{ds}$ and $\frac{dy}{ds}$ both exist and are not simultaneously null is zero.

Proof. After placing the curve in the first quadrant at a positive distance from the origin, parametrize it according to arc length, in Cartesian coordinates

$$x = x(s), \qquad y = y(s), \qquad 0 \leqq s \leqq s_0,$$

[7] $m_C(A) = {}_{df}$ Carathéodory linear measure of A. For definition and properties, see [3] and [9]—especially pp. 53–54.

[8] $|\,B\,|_i = {}_{df}$ Lebesgue i-dimensional measure of B.

and in polar coordinates

$$R = R(s), \qquad \alpha = \alpha(s), \qquad 0 \leqq s \leqq s_0 .$$

We note that $x(s)$, $y(s)$, $R(s)$, and $\alpha(s)$ all satisfy Lipschitz conditions of order 1 and so are absolutely continuous. Also $x(s) > 0$, $y(s) > 0$, $M > R(s) > 0$, $\frac{1}{2}\pi > \alpha(s) > 0$. Therefore, if θ is fixed,

$$R(s) \cos [\alpha(s) - \theta] \tag{9}$$

satisfies a Lipschitz condition of order 1. By Saks, *Theory of the Integral*, Chapter IV, (8.4), (iii), $\frac{dx}{ds}$ and $\frac{dy}{ds}$ exist for almost all s. From the measurability of $\frac{dx}{ds}$ and $\frac{dy}{ds}$, the absolute continuity of $x(s)$ and $y(s)$, and the second part of Saks (8.4), (iv), $\frac{dx}{ds}$ and $\frac{dy}{ds}$ are simultaneously zero on a set of s-values of measure zero. Thus the first part of our lemma is proved.

Since $\arctan\left(\frac{dy}{ds} \Big/ \frac{dx}{ds}\right)$, $\frac{dx}{ds} \neq 0$, is a measurable function of s it is easy to see that $| \underset{(\theta,s)}{E} [\theta = \theta_s] |_2 = 0$, where θ_s is the θ corresponding to the line tangent to the curve at $(x(s), y(s))$. Consider now the transformation

$$\begin{aligned} \theta &= \theta, \\ \rho &= R(s) \cos [\alpha(s) - \theta]. \end{aligned} \tag{10}$$

This transformation is continuous. Cover $E = \underset{(\theta,s)}{E} [\theta = \theta_s]$ by an open set $E(\epsilon)$ such that $| E(\epsilon) |_2 < \epsilon$. $E(\epsilon)$ is open and so is an[9] $\mathcal{F}_\sigma$. Its transform $\tau(E(\epsilon))$ in the (ρ, θ)-plane is an[10] $\mathcal{F}_\sigma$ and so is measurable Lebesgue. By the Fubini theorem

$$| \tau(E(\epsilon)) |_2 = \int \left[\int C_{\tau(E(\epsilon))}(\rho, \theta)\, d\rho \right] d\theta.$$

Since $R(s) \cos [\alpha(s) - \theta]$ satisfies a Lipschitz condition of order 1, there exists a constant M depending only on the curve such that

$$\int C_{\tau(E(\epsilon))}(\rho, \theta)\, d\rho \leqq M \int C_{E(\epsilon)}(\theta, s)\, ds,$$

and so

$$| \tau(E(\epsilon)) |_2 \leqq M | E(\epsilon) |_2 \leqq M\epsilon.$$

Thus $| \tau(E) |_2 = 0$ and the second part of our lemma is proved.

[9] See [11], Theorem 52a.
[10] See [11], Theorem 41.

THEOREM 1. *If $\mathcal{C}$ is a continuous rectifiable curve, then*

$$St(\mathcal{C}) = 2l(\mathcal{C}).$$

Proof. Parametrize $\mathcal{C}$ according to arc length. Consider a sequence of approximating polygons (with vertices on the curve) formed by interpolating more and more vertices, where the maximum side approaches zero in length, and where l_n (the length of the n-th approximating polygon) approaches l (the length of the curve) as $n \to \infty$. If $\varphi_n(\rho, \theta)$ is the number of times (calculated with the proper multiplicity) that (ρ, θ) intersects the n-th polygon, then

$$\varphi_{n+1}(\rho, \theta) \geqq \varphi_n(\rho, \theta),$$

except for at most a finite set of lines (the sides of the n-th polygon). We now show that for a sequence satisfying the above conditions

$$\lim_{n\to\infty} \varphi_n(\rho, \theta) = f_{\mathcal{C}}(\rho, \theta), \tag{11}$$

except for at most a set of lines tangent to the curve through points of the curve where either $\frac{dx}{ds}$ fails to exist, $\frac{dy}{ds}$ fails to exist, or $\frac{dx}{ds} = \frac{dy}{ds} = 0$. As can be seen from Lemmas I, II, and Saks (8.4), (ii), the Deltheil measure of this set of lines is zero.

We wish to show that, except for the set of lines of Deltheil measure zero, equation (11) holds. Suppose (ρ_0, θ_0) is a line not in the set of measure zero. If $f_{\mathcal{C}}(\rho_0, \theta_0) = 0$, it is not difficult to see that $\varphi_n(\rho_0, \theta_0) = 0$. If $f_{\mathcal{C}}(\rho_0, \theta_0) = K$ $(0 < K < \infty)$, then we can find an increasing (or decreasing) sequence of different s_i's, such that $(x(s_i), y(s_i))$, $1 \leqq i \leqq K$, is on line (ρ_0, θ_0). If $f_{\mathcal{C}}(\rho_0, \theta_0) = \infty$, we can find by the procedure generally used to prove the Bolzano-Weierstrass theorem an increasing (or decreasing) sequence of different s_i's, such that again $(x(s_i), y(s_i))$, $1 \leqq i < \infty$, is on the line (ρ_0, θ_0). If there exist s_i' and s_i'' such that $0 \leqq s_{i-1} < s_i' \leqq s_i \leqq s_i'' < s_{i+1} \leqq l(\mathcal{C})$ and s_i' and s_i'' determine succeeding vertices in the n-th polygon, then we say that s_i is *effective* in $\varphi_n(\rho_0, \theta_0)$. It is obvious that if s_i is effective in $\varphi_N(\rho_0, \theta_0)$, then s_i is effective in each $\varphi_n(\rho_0, \theta_0)$, $N \leqq n$. For n large enough s_j is effective in $\varphi_n(\rho_0, \theta_0)$. Let us suppose that $\left(\frac{dx}{ds}\right)_{s=s_j} > 0$ and $\left(\frac{dy}{ds}\right)_{s=s_j} > 0$. Then there exists an η such that

$$y(s_{j-1}) < y(s) < y(s_j), \quad x(s_{j-1}) < x(s) < x(s_j), \quad s_j - \eta \leqq s < s_j; \tag{12}$$

$$y(s_j) < y(s) < y(s_{j+1}), \quad x(s_j) < x(s) < x(s_{j+1}), \quad s_j \leqq s < s_j + \eta; \tag{13}$$

and

$$\left|\frac{y(s) - y(s_j)}{x(s) - x(s_j)} - \left(\frac{dy}{dx}\right)_{s=s_j}\right| < \min\left[\frac{1}{2}\left|\tan(\theta_0 + \tfrac{1}{2}\pi) - \left(\frac{dy}{dx}\right)_{s=s_j}\right|, \frac{1}{2}\left|\left(\frac{dy}{dx}\right)_{s=s_j}\right|\right], \quad 0 < |s_j - s| < \eta. \tag{14}$$

This means that, if s lies in the range described by (14), the points will be in a sector with $(x(s_j), y(s_j))$ as center such that (ρ_0, θ_0) is not in the sector and such that the upper half of the sector will be above the horizontal line through the center and to the right of the vertical line through the center. The range described in (12) has points only in the upper half-sector and so on one side of the line (ρ_0, θ_0); the range described in (13) has points only in the lower half-sector and so on the other side of the line (ρ_0, θ_0). For n large enough, say N', the maximum arc length between successive vertices is less than η, and s_j will be effective in $\varphi_{N'}(\rho_0, \theta_0)$. If s_j does not determine a vertex of the n-th polygon, then a pair of succeeding vertices will be in opposite half-sectors, and by a topological argument the side joining them will be intersected by (ρ_0, θ_0). Thus after disposing of the other cases, e.g., $\frac{dy}{ds} = 0, \frac{dx}{ds} < 0$, we see that s_j "contributes to" $\varphi_n(\rho_0, \theta_0)$ for n large enough. And by induction

$$\lim_{n\to\infty} \varphi_n(\rho_0, \theta_0) = f_{\mathcal{C}}(\rho_0, \theta_0).$$

But

$$2l(\mathcal{C}) = \lim_{n\to\infty} 2l_n = \lim_{n\to\infty} \iint \varphi_n(\rho, \theta)\, d\rho\, d\theta$$

$$= \iint \lim_{n\to\infty} \varphi_n(\rho, \theta)\, d\rho\, d\theta = \iint f_{\mathcal{C}}(\rho, \theta)\, d\rho\, d\theta = St(\mathcal{C}),$$

and the theorem is proved.

COROLLARY. *If $\mathcal{C}$ is a rectifiable Jordan arc, then $St(C) = St(\mathcal{C}) = 2l(\mathcal{C}) = 2m_c(C)$, where C is the corresponding point-set.*

LEMMA 3. *If A is closed and bounded, then $T(A)$ is closed and bounded.*

Proof. Since A is bounded, for $(\rho, \theta) \in T(A)$, ρ is bounded and, of course, θ is bounded $(2\pi \geqq \theta \geqq 0)$. Now that we have $T(A)$ bounded, let us prove that $T(A)$ is closed. Let $\{(\rho_i, \theta_i)\}$ be a convergent sequence of different lines of $T(A)$. We must prove that $(\rho_i, \theta_i) \to (\rho_0, \theta_0) \in T(A)$ as $i \to \infty$. For each (ρ_i, θ_i) consider one of its inverses $(R_i, \alpha_i) \in A$. Either the set $\{(R_i, \alpha_i)\}$ has only a finite set of distinct elements or it has an infinite set of distinct elements. In the first case there would be an infinite number of distinct elements of $\{(\rho_i, \theta_i)\}$ on a closed bounded curve in $T(A)$, and $(\rho_i, \theta_i) \to (\rho_0, \theta_0) \in T(A)$. In the second case since A is self-compact, there is a subsequence $(R_i', \alpha_i') \to (R_0, \alpha_0) \in A$. This convergent subsequence determines a subsequence $\{(\rho_i', \theta_i')\}$. But $\lim_{i\to\infty} (\rho_i', \theta_i') = (R_i' \cos(\alpha_i' - \theta_i'), \theta_i') = (R_0 \cos(\alpha_0 - \theta_0), \theta_0) \subset E_{(\rho,\theta)}[\rho = R_0 \cos(\alpha_0 - \theta), \rho \geqq 0] = T[(R_0, \alpha_0)] \subset T(A)$. Hence $\lim_{i\to\infty} (\rho_i, \theta_i) \in T(A)$ and $T(A)$ is closed.

LEMMA 4. *If A is an $\mathcal{F}_\sigma$ in the (R, α)-plane, then $T(A)$ is an $\mathcal{F}_\sigma$ in the (ρ, θ)-plane.*

LEMMA 5. *If $m_e A = a < \infty$, then $T(A)$ is measurable Lebesgue.*

Proof. Since $m_e A < \infty$, we have $A = F + 0$, where F is an $\mathcal{F}_\sigma$ in the (R, α)-plane and $m_e(0) = 0$. Hence $T(F)$ is an $\mathcal{F}_\sigma$ in the (ρ, θ)-plane. Since $m_e(0) = 0$, $| T(0) |_2 = 0$. But $T(A) = T(F + 0) = T(F) + T(0)$ and so $T(A)$ is measurable Lebesgue.

LEMMA 6 [2]. *If $m_e(A) < \infty$, A is irregular, and A_θ is the projection of A on line (ρ, θ), then $m_e(A_\theta) = 0$ for almost all θ.*

LEMMA 6.1. *If $m_e A < \infty$ and A is irregular, then $St(A) = 0$.*

Proof. By Lemma 5, $C_{T(A)}(\rho, \theta)$ is measurable and by the Fubini Theorem

$$| T(A) |_2 = \int\int C_{T(A)}(\rho, \theta)\, d\rho\, d\theta = \int \left[\int C_{T(A)}(\rho, \theta)\, d\rho \right] d\theta.$$

By Lemma 6

$$\int C_{T(A)}(\rho, \theta)\, d\rho = 0$$

for almost all θ. Hence

$$| T(A) |_2 = 0$$

and

$$St(A) = 0.$$

LEMMA 7. *If a plane curve $\mathcal{C}$ is rectifiable, then there exists a continuous rectifiable curve $\mathcal{C}'$ which is of equal length and contains $\mathcal{C}$ as a subset.*

Proof. Let $\mathcal{C}$ be given by the equations

$$x = X(t), \qquad y = Y(t), \qquad 0 \leqq t \leqq 1.$$

$S(\mathcal{C}; 0, t)$ is a monotone, single-valued function of t, $0 \leqq t \leqq 1$. If this function is continuous, then $\mathcal{C}$ is continuous. If the function is not continuous and t_1 is a point of discontinuity, let $S(t_1^-) = \sup S(\mathcal{C}; 0, t)$ and $S(t_1^+) = \inf S(\mathcal{C}; 0, t)$. We have $S(t_1^-) < S(t_1^+)$. Now parametrize $\mathcal{C}$ in terms of arc length so that each point is given by $(x(s), y(s))$ for some s, $0 \leqq s \leqq s_0$. Since $S(\mathcal{C}; 0, t)$ is not continuous, every such s does not have a corresponding point. Consider $\mathcal{C}'$ as the curve generated by adding to $\mathcal{C}$ those parts of the line segments $(x(s^-), y(s^-))$, $(x(s^+), y(s^+))$ not already contained in the curve. This new curve has the same length as $\mathcal{C}$. When it is parametrized according to arc length, values of s which yielded points on $\mathcal{C}$ yield the same points on $\mathcal{C}'$, and values of s which failed to yield points on $\mathcal{C}$ yield points on the new line segment.

LEMMA 8. *Any regular set is almost entirely contained in a countable class of rectifiable Jordan arcs.*

Proof. By [1], Theorem 16 and the preceding lemma, we show that any regular set is contained in a set of measure zero and a countable class of recti-

fiable arcs. By a modification of [1], Lemma 4 (change (i) to read: G is a set of rectifiable Jordan arcs), the rest of the theorem follows.

LEMMA 9. *If $m_c(A) < \infty$ and J is a rectifiable Jordan arc, then $2m_c(AJ) = St(AJ)$.*

Proof follows immediately from Corollary to Theorem 1 and the complete additivity of Steinhaus measure.

THEOREM 3. *If A is regular and $m_c(A) < \infty$, then $St(A) = 2m_cA$.*

Proof follows from Lemma 8, Lemma 9, and the complete additivity of Steinhaus measure.

THEOREM 4. *If $m_c(A) < \infty$ or if $A = \sum_{i=1}^{\infty} A_i$, $m_c(A_i) < \infty$, and A_R is the regular part of A, then $St(A) = 2m_c(A_R)$.*

Proof. If $m_c(A) < \infty$, then $A = A_I + A_R$, where A_I is the irregular part of A and A_R is the regular part of A. By Theorems 2 and 3,

$$St(A) = St(A_I) + St(A_R) = 0 + St(A_R) = 2m_c(A_R).$$

The other case offers no new difficulties.

Multiple points. Let $\mathcal{C}$ be a continuous rectifiable curve parametrized according to arc length. Let E_i be that set of points which have exactly i correspondents in the s-range.

LEMMA 10. *For $i \leqq \infty$, E_i is measurable.*

Proof. Let $E_i(j) = {}_{d/p}E$. $\Big[$There exists a set of numbers $\{\tau_K(p)\}_{1 \leqq K \leqq i}$ such that $\tau_K(p) < \tau_{K'}(p), K < K'; (x(\tau_K(p)), y(\tau_K(p))) = p; \min_{1 \leqq K < i} | \tau_{K+1}(p) - \tau_K(p) | \geqq \frac{1}{j}\Big]$. Corresponding to each $p \in E_i(j)$, there may be many sets which satisfy the conditions above, but we associate with each p a unique set which satisfies the prescribed conditions.

We now prove that $E_i(j)$ is closed. For each sequence $\{p_n\}$ such that $p_n \in E_i(j)$, $p_n \to p$, we can choose a subsequence $\{p_{n'}\}$ such that $\tau_K(p_{n'}) \to a_K$, $1 \leqq K \leqq i$. Since for each n', $\min_{1 \leqq K < i} | \tau_{K+1}(p_{n'}) - \tau_K(p_{n'}) | \geqq \frac{1}{j}$, we have $\min_{1 \leqq K < i} | a_{K+1} - a_K | \geqq \frac{1}{j}$. Thus $p \in E_i(j)$ and $E_i(j)$ is closed.

But

$$\sum_{i \leqq K \leqq \infty} E_K = \sum_{j=1}^{\infty} E_i(j) \in \mathcal{F}_\sigma.$$

Since for $i < \infty$, $E_i = \sum_{i \leqq K \leqq \infty} E_K - \sum_{i+1 \leqq K \leqq \infty} E_K$ is the difference between two meas-

urable sets, we have E_i measurable. Since $E_\infty = \sum_{1 \leqq i < \infty} E_i(j) - \sum_{i \leqq K < \infty} E_K$, we also have that E_∞ is measurable.

THEOREM 5. *If $\mathcal{C}$ is a continuous rectifiable curve parametrized according to arc length, then*

$$l(\mathcal{C}) = m_c(E_i).$$

Proof. Let $\mathfrak{S}_i$ be the set of all points of E_i counted with proper multiplicity. Then $2l(\mathcal{C}) = St(\mathcal{C}) = \iint f_{\mathcal{C}}(\rho, \theta)\, d\rho\, d\theta = \sum_{1 \leqq i \leqq \infty} \iint f_{\mathfrak{S}_i}(\rho, \theta)\, d\rho\, d\theta =$

$$\sum_{1 \leqq i \leqq \infty} i \iint f_{E_i}(\rho, \theta)\, d\rho\, d\theta = \sum_{1 \leqq i < \infty} iSt(E_i) = 2 \sum_{1 \leqq i < \infty} im_c(E_i).$$

The last theorem is a sharpening of a remark made by Saks. It would seem that a more direct (i.e., without reverting to Steinhaus measure) proof should be possible but the author has not been able to find such a proof.

BIBLIOGRAPHY

1. A. S. BESICOVITCH, *On the fundamental geometrical properties of linearly measurable sets of points* (II), Math. Annalen, vol. 115(1938), pp. 296–329.
2. A. S. BESICOVITCH, *On the fundamental geometrical properties of linearly measurable sets of points* (III), Math. Annalen, vol. 116(1939), pp. 349–357.
3. C. CARATHÉODORY, *Über das lineare Mass von Punktmengen—eine Verallgemeinerung des Längenbegriffs*, Nachr. Ges. Wiss. Göttingen, 1914, pp. 404–426.
4. J. FAVARD, *Une définition de la longeur et de l'aire*, Comptes Rendus, vol. 194(1932), pp. 344–346.
5. W. GROSS, *Über das Flächenmass von Punktmengen*, Monatshefte für Mathematik und Physik, vol. 29(1918), pp. 145–176.
6. A. KOLMOGOROFF, *Beiträge zur Masstheorie*, Math. Annalen, vol. 107(1932), pp. 351–366.
7. J. F. RANDOLPH, *On generalizations of length and area*, Bull. A. M. S., vol. 42(1936), pp. 268–274.
8. J. RANDOLPH AND A. MORSE, *Gillespie measure*, Duke Mathematical Journal, vol. 6 (1940), pp. 408–419.
9. S. SAKS, *Theory of the Integral*, second revised edition, Warsaw-Lwów, 1937.
10. E. SCHMIDT, *Über die definition des Begriffs der Länge krummer Linien*, Math. Annalen, vol. 55(1902), pp. 163–176.
11. W. SIERPIŃSKI, *Introduction to General Topology*, Toronto, 1934.
12. H. STEINHAUS, *Zur Praxis der Rektifikation und zum Längenbegriff*, Sächsiche Akademie zu Leipzig, Berichte, vol. 82(1930), pp. 120–130.
13. H. STEINHAUS, *Sur la portée pratique et théorique de quelques théorèmes sur la mesure des ensembles de droites*, Comptes Rendus du premier congrès des mathématiciens des pays Slaves, 1929, pp. 348–354.

CORNELL UNIVERSITY.

LIMITS OF INTEGRALS

By Ralph Palmer Agnew

1. **Introduction.** Let integrals over finite intervals be Lebesgue integrals, and let integrals over infinite intervals be Cauchy-Lebesgue integrals defined by

$$\int_0^\infty = \lim_{A\to\infty} \int_0^A ; \qquad \int_{-\infty}^\infty = \lim_{A,B\to\infty} \int_A^B . \tag{1.1}$$

In case $g(t)$ is integrable over each finite interval, the identity

$$\int_{A-\lambda}^{B} [g(t+\lambda) - g(t)]\, dt = \int_B^{B+\lambda} g(t)\, dt - \int_A^{A+\lambda} g(-t)\, dt \tag{1.2}$$

implies that the equality

$$\int_{-\infty}^{\infty} [g(t+\lambda) - g(t)]\, dt = \lim_{A\to\infty} \int_A^{A+\lambda} g(t)\, dt - \lim_{A\to\infty} \int_A^{A+\lambda} g(-t)\, dt \tag{1.3}$$

holds whenever either of the two members exists. For this and other reasons, facts relating to

$$\lim_{A\to\infty} \int_A^{A+\lambda} f(t)\, dt \tag{1.4}$$

are of interest.

Perhaps our most striking result is that if the limit in (1.4) exists for each λ in some set having positive measure, then the limit exists for each real λ and the convergence is *uniform* over each finite interval. Some applications of this result are given in §4 and §5.

2. **A preliminary theorem.** Our first step in the study of (1.4) is to prove the following theorem.

Theorem 2.1. *If*

$$L(\lambda) = \lim_{A\to\infty} \int_A^{A+\lambda} f(t)\, dt \tag{2.2}$$

exists for each λ in some set having positive measure, then $L(\lambda)$ exists for each real λ and $L(\lambda) = \lambda L$ where $L = L(1)$.

The hypothesis implies existence of a number a such that $f(t)$ is integrable over $b \leqq t \leqq c$ provided b and c are greater than a; all limits of integration which we use are assumed to be greater than a. The identity

$$\int_A^{A+\lambda_2-\lambda_1} = \int_A^{A+\lambda_2} - \int_{(A+\lambda_2-\lambda_1)}^{(A+\lambda_2-\lambda_1)+\lambda_1} , \tag{2.3}$$

Received May 10, 1941; presented to the American Mathematical Society, May 2, 1941.

with integrand $f(t)$, implies that $L(\lambda_2 - \lambda_1)$ exists and

$$L(\lambda_2 - \lambda_1) = L(\lambda_2) - L(\lambda_1) \tag{2.4}$$

whenever $L(\lambda_1)$ and $L(\lambda_2)$ both exist. Since the set E of values of λ for which $L(\lambda)$ exists has positive measure, there is a positive number δ such that each number λ_0 for which $|\lambda_0| < \delta$ is representable in the form $\lambda_0 = \lambda_2 - \lambda_1$ where λ_2 and λ_1 are points of E.[1] Hence $L(\lambda)$ exists when $|\lambda| < \delta$. The identity

$$\int_A^{A+\lambda_1+\lambda_2} = \int_A^{A+\lambda_1} + \int_{(A+\lambda_1)}^{(A+\lambda_1)+\lambda_2}, \tag{2.5}$$

with integrand $f(t)$, implies that $L(\lambda_1 + \lambda_2)$ exists and

$$L(\lambda_1 + \lambda_2) = L(\lambda_1) + L(\lambda_2) \tag{2.6}$$

provided $L(\lambda_1)$ and $L(\lambda_2)$ both exist. It is now easy to show that $L(\lambda)$ exists for each real λ, and that (2.4) and (2.6) hold whenever λ_1 and λ_2 are real. From (2.4) we see that $L(\lambda)$ is continuous everywhere or discontinuous everywhere according as $L(\lambda)$ is continuous or discontinuous at $\lambda = 0$. Since $L(\lambda)$, being the limit of the sequence of continuous functions obtained by giving integer values to A, cannot be discontinuous everywhere it must be continuous everywhere. The functional equation (2.5) implies, in a familiar and simple way, that $L(r) = rL(1)$ when r is rational; continuity of $L(\lambda)$ then implies that $L(\lambda) = \lambda L(1)$ for each λ and Theorem 2.1 is proved.

3. **Uniformity of the convergence.** In this section, we prove the following theorem.

THEOREM 3.1. *If*

$$\lim_{A\to\infty} \int_A^{A+\lambda} f(t)\,dt = \lambda L, \qquad -\infty < \lambda < \infty, \tag{3.2}$$

then the convergence in (3.2) *is uniform over each finite interval* $-a \leqq \lambda \leqq a$.

Let a be a fixed positive number, let E_1 denote the interval $-a \leqq \lambda \leqq a$, and let the measure of E_1 be denoted by $|E_1|$ so that $|E_1| = 2a$. By a theorem of Egoroff, the convergence in (3.2) is essentially uniform over E_1; that is, to each $\theta > 0$ there corresponds a subset E of E_1 such that $|E| > |E_1| - \theta$ and the convergence is uniform over E. Let θ and E be fixed such that $\theta < \frac{1}{2}a$ and accordingly $|E| > \frac{3}{2}a$. Let $A(\epsilon)$ be a function, defined for $\epsilon > 0$, such that

$$\left| \int_A^{A+\lambda} f(t)\,dt - \lambda L \right| < \tfrac{1}{2}\epsilon, \qquad \lambda \in E,\ A > A(\epsilon). \tag{3.3}$$

[1] This fact, first proved by Steinhaus, Fundamenta Mathematicae, vol. 1(1920), pp. 93–104, has since received very simple proofs. One chooses a point at which the density is greater than $\frac{1}{2}$ and applies the idea following equation (3.4) below. It is an interesting fact that some sets of measure 0, notably the Cantor middle-third set, have the essential property which we are using.

Let λ_0 represent any point in the interval $-a \leqq \lambda \leqq a$. Then points λ_2 and λ_1 of the set E exist such that

$$\lambda_0 = \lambda_2 - \lambda_1 . \tag{3.4}$$

To prove this, we observe that if such a representation of λ_0 were impossible, then the set E could have no points in common with the set E_0 obtained by translating the set E to the right $|\lambda_0|$ units; this would lead to the absurd conclusion that E and E_0 are two disjoint subsets of the interval $-a \leqq \lambda \leqq 2a$ each having measure greater than $\frac{3}{2}a$.

Use of the representation (3.4) and the inequality (3.3) gives, when $A > a + A(\epsilon)$ and $-a \leqq \lambda_0 \leqq a$,

$$\left| \int_A^{A+\lambda_0} f(t)\,dt - \lambda_0 L \right| = \left| \int_A^{A+\lambda_2-\lambda_1} f(t)\,dt - (\lambda_2 - \lambda_1)L \right|$$

$$= \left| \left\{ \int_A^{A+\lambda_2} f(t)\,dt - \lambda_2 L \right\} - \left\{ \int_{(A+\lambda_2-\lambda_1)}^{(A+\lambda_2-\lambda_1)+\lambda_1} f(t)\,dt - \lambda_1 L \right\} \right| < \epsilon$$

and the uniform convergence is established.

Because of the identity (1.2), it is a consequence of Theorems 2.1 and 3.1 that if

$$\lim_{A,B\to\infty} \int_A^B [g(t+\lambda) - g(t)]\,dt$$

exists for each λ in some set having positive measure, then there is a constant M such that the limit is λM *uniformly over each finite interval of values of* λ.

4. **Two theorems of Iyengar.** In this section we prove two theorems which, as may be seen by making an exponential change of variable and suitable changes in notation, imply and are implied by a theorem and other results of Iyengar.[2] The proofs of Iyengar are ingenious; by making use of Theorem 3.1 we obtain simpler proofs.[3]

THEOREM 4.1. *A necessary and sufficient condition that*

$$\lim_{A\to\infty} \int_A^{A+\lambda} f(t)\,dt \tag{4.11}$$

[2] K. S. K. Iyengar, *On Frullani integrals*, Proc. Cambridge Philos. Soc., vol. 37(1941), pp. 9-13. The condition

$$\lim_{\epsilon\to 0} \int_{\epsilon\rho}^{\epsilon} \frac{\varphi(u)}{u}\,du = B \log \rho, \qquad \rho > 0,$$

of Iyengar becomes (4.21) when we set

$$t = \log u^{-1},\ A = \log \epsilon^{-1},\ \lambda = \log \rho^{-1},\ L = -B,\ f(t) = \varphi(e^{-t}).$$

[3] The author must confess that these theorems seem strange to him; he and some of his colleagues feel it to be incredible that no Tauberian conditions are involved in the theorems.

exist for each real λ *is that*

$$\lim_{A\to\infty} e^{A}\int_{A}^{\infty} f(t)e^{-t}\,dt \tag{4.12}$$

exist.

THEOREM 4.2. *A necessary and sufficient condition that*

$$\lim_{A\to\infty}\int_{A}^{A+\lambda} f(t)\,dt = \lambda L, \qquad -\infty < \lambda < \infty, \tag{4.21}$$

is that

$$\lim_{A\to\infty} e^{A}\int_{A}^{\infty} f(t)e^{-t}\,dt = L. \tag{4.22}$$

Because of the equality

$$\begin{aligned} e^{-A}\int_{A}^{\infty} f(t)e^{-t}\,dt &= \int_{A}^{\infty} f(t)e^{-(t-A)}\,dt \\ &= \int_{0}^{\infty} f(t+A)e^{-t}\,dt, \end{aligned} \tag{4.23}$$

which holds whenever any one of the integrals exists, the conditions (4.12) and (4.22) can be put in different forms.

Using Theorem 2.1, we can see that Theorem 4.1 is a corollary of Theorem 4.2. We prove Theorem 4.2. To prove necessity, let a be a fixed positive number; we could take $a = 1$. Then, by Theorem 3.1, to each $\epsilon > 0$ corresponds a number $A_0 \equiv A_0(\epsilon) > 0$ such that

$$\left|\int_{A}^{A+\lambda} f(t)\,dt - \lambda L\right| < \tfrac{1}{2}\epsilon(1 - e^{-a}), \quad A > A_0,\ 0 \leqq \lambda \leqq a. \tag{4.24}$$

Let

$$f_1(t) = f(t) - L, \tag{4.25}$$

so that (4.24) may be written

$$\left|\int_{A}^{A+\lambda} f_1(t)\,dt\right| < \tfrac{1}{2}\epsilon(1 - e^{-a}), \qquad A > A_0,\ 0 \leqq \lambda \leqq a. \tag{4.26}$$

Let x and A be momentarily fixed such that $x > A > A_0$, and choose an index N such that

$$A + Na < x \leqq A + (N+1)a. \tag{4.27}$$

We are going to use the inequality

$$\left|\int_{A}^{x}\right| \leqq \left|\int_{A}^{A+a}\right| + \left|\int_{A+a}^{A+2a}\right| + \cdots + \left|\int_{A+(N-1)a}^{A+Na}\right| + \left|\int_{A+Na}^{x}\right| \tag{4.28}$$

with integrand $e^{A}f_1(t)e^{-t}$. Using the second mean value theorem[4] we obtain for each $n = 1, 2, \cdots, N$

$$(4.29)\quad \int_{A+(n-1)a}^{A+na} f_1(t)e^{-t}\,dt = e^{-A-(n-1)a}\int_{A+(n-1)a}^{\xi_n} f_1(t)\,dt + e^{-A-na}\int_{\xi_n}^{A+na} f_1(t)\,dt,$$

where ξ_n is a properly chosen point between $A + (n-1)a$ and $A + na$. Using (4.29) and (4.26) we obtain

$$(4.30)\quad \left| e^{A}\int_{A+(n-1)a}^{A+na} f_1(t)e^{-t}\,dt \right| < \epsilon(1 - e^{-a})e^{-(n-1)a}.$$

Likewise

$$(4.31)\quad \left| e^{A}\int_{A+Na}^{x} f_1(t)e^{-t}\,dt \right| < \epsilon(1 - e^{-a})e^{-Na}.$$

From (4.28), (4.30), and (4.31) we obtain

$$(4.32)\quad \left| e^{A}\int_{A}^{x} f_1(t)e^{-t}\,dt \right| < \epsilon, \qquad x > A > A_0(\epsilon).$$

Since $A_0(\epsilon)$ was chosen greater than 0, this implies that

$$(4.33)\quad \left| \int_{A}^{x} f_1(t)e^{-t}\,dt \right| < \epsilon, \qquad x > A > A_0(\epsilon);$$

and hence the Cauchy criterion for convergence implies existence of

$$(4.34)\quad \int_{B}^{\infty} f_1(t)e^{-t}\,dt$$

for each sufficiently great constant B. Hence we can let x become infinite in (4.32) to obtain

$$(4.35)\quad \left| e^{A}\int_{A}^{\infty} f_1(t)e^{-t}\,dt \right| \leqq \epsilon, \qquad A > A_0(\epsilon),$$

and therefore

$$(4.36)\quad \left| e^{A}\int_{A}^{\infty} f(t)e^{-t}\,dt - L \right| \leqq \epsilon, \qquad A > A_0(\epsilon).$$

This completes proof of necessity for Theorem 4.2.

To prove sufficiency, let

$$(4.37)\quad F(t) = e^{t}\int_{t}^{\infty} f(u)e^{-u}\,du;$$

[4] In case $f(t)$ is complex valued, we obtain the results separately for the real and imaginary parts of $f(t)$ and Theorem 4.2 then follows.

our hypothesis (4.22) then becomes

$$\lim_{t\to\infty} F(t) = L. \tag{4.38}$$

Differentiating (4.37) gives for almost all t (that is, for all t except those in a set of measure 0)

$$F'(t) = F(t) - f(t). \tag{4.39}$$

Integrating over the interval with end points at A and $A + \lambda$ gives

$$F(A + \lambda) - F(A) = \int_A^{A+\lambda} F(t)\,dt - \int_A^{A+\lambda} f(t)\,dt. \tag{4.40}$$

Because of (4.38), we can let A become infinite in (4.40) to obtain (4.21). This completes proof of Theorems 4.1 and 4.2.

5. **A Tauberian theorem.** We now use Theorem 4.2 to prove the following theorem which could be easily generalized by replacing $F(t)$ by $F(t) - L$, L being a constant.

THEOREM 5.1. *If $F(t)$ is absolutely continuous over each finite interval and*

$$\lim_{A\to\infty} \int_A^{A+\lambda} [F(t) - F'(t)]\,dt = 0, \qquad -\infty < \lambda < \infty, \tag{5.11}$$

and

$$\lim_{t\to\infty} e^{-t}F(t) = 0, \tag{5.12}$$

then

$$\lim_{t\to\infty} F(t) = 0. \tag{5.13}$$

To prove this theorem, put

$$f(t) = F(t) - F'(t). \tag{5.14}$$

Then, by Theorem 4.2,

$$\lim_{t\to\infty} e^{t} \int_t^{\infty} f(u)e^{-u}\,du = 0. \tag{5.15}$$

Multiplying (5.14) by the integrating factor e^{-t} gives

$$\frac{d}{dt}\,e^{-t}F(t) = -f(t)e^{-t} \tag{5.16}$$

and, since (5.15) implies existence of the integral on the right,

$$e^{-t}F(t) = c + \int_t^{\infty} f(u)e^{-u}\,du. \tag{5.17}$$

The result of letting t become infinite shows that $c = 0$. Hence

$$F(t) = e^t \int_t^\infty f(u) e^{-u}\, du \tag{5.18}$$

and our result follows from (5.15).

That the Tauberian condition (5.12) cannot be removed from Theorem 5.1 is an obvious consequence of the fact that the function $F(t) = e^t$ satisfies (5.11) but not (5.13).

6. **Bounded functions $f(t)$.** It was pointed out to the author by R. P. Boas that, in case $f(t)$ is bounded and measurable and the equality

$$\lim_{A\to\infty} \int_A^{A+\lambda} f(t)\, dt = \lambda L, \tag{6.01}$$

in which L is a constant, holds for two values λ_1 and λ_2 of λ for which λ_1/λ_2 is irrational, an application of a Tauberian theorem of Wiener establishes the equality

$$\lim_{A\to\infty} e^A \int_A^\infty f(t) e^{-t}\, dt = L; \tag{6.02}$$

in this case (6.01) holds also for all values of λ. It is a consequence of a Tauberian theorem of Wiener[5] that if $K_1(t)$, $K_2(t)$, $K_3(t)$ are three functions having absolutely convergent integrals over $-\infty < t < \infty$, if

$$\int_{-\infty}^\infty K_j(t)\, dt = 1, \qquad j = 1, 2, 3, \tag{6.11}$$

if the Fourier transforms of $K_1(t)$ and $K_2(t)$ have no common zeros, and if $f(t)$ is a bounded measurable function for which

$$\lim_{A\to\infty} \int_{-\infty}^\infty K_j(t - A) f(t)\, dt = L \tag{6.12}$$

when $j = 1, 2$, then (6.12) holds when $j = 3$. On setting, when $j = 1, 2$,

$$K_j(t) = \lambda_j^{-1} \quad \text{for } 0 \leqq t \leqq \lambda_j,$$
$$K_j(t) = 0 \quad \text{otherwise,}$$

and

$$K_3(t) = e^{-t} \quad \text{for } t \geqq 0,$$
$$K_3(t) = 0 \quad \text{for } t < 0, \tag{6.13}$$

we obtain the conclusion (6.02). If we set, for $\lambda > 0$,

$$K_3(t) = \lambda^{-1} \quad \text{for } 0 \leqq t \leqq \lambda,$$
$$K_3(t) = 1 \quad \text{otherwise,}$$

[5] N. Wiener, *The Fourier Integral*, Cambridge (1933), p. 75, Theorem 6.

we obtain (6.01) for $\lambda > 0$; and it then follows easily that (6.01) holds when $\lambda \leqq 0$. That (6.02) implies (6.01) follows, in case $f(x)$ is bounded, from Theorem 4, p. 73, of Wiener's book and the fact that the Fourier transform of the function $K_3(t)$ in (6.13) has no zeros.

Even when $f(t)$ is not assumed bounded, the hypothesis that (6.01) holds when $\lambda = \lambda_1$ and when $\lambda = \lambda_2$ is no less general than the hypothesis that the left member of (6.01) exists when $\lambda = \lambda_1$ and when $\lambda = \lambda_2$. This is a consequence of the following theorem.

THEOREM 6.2. *If*

$$\lim_{A\to\infty} \int_A^{A+\lambda} f(t)\,dt \tag{6.21}$$

exists for some $\lambda \neq 0$, *and* t_0 *is fixed such that* $f(t)$ *is integrable over each finite interval* $t_0 \leqq t \leqq B < \infty$, *then*

$$\lim_{B\to\infty} \frac{1}{B} \int_{t_0}^{B} f(t)\,dt \tag{6.22}$$

exists; moreover the equality

$$\lim_{A\to\infty} \int_A^{A+\lambda} f(t)\,dt = \lambda \lim_{B\to\infty} \frac{1}{B} \int_{t_0}^{B} f(t)\,dt \tag{6.23}$$

holds for each λ *for which the left member exists.*

Assuming first that λ is a positive number for which (6.21) exists, we prove (6.23). On denoting the limit by λL and setting $f_1(t) = f(t) - L$, we obtain

$$\lim_{A\to\infty} \int_A^{A+\lambda} f_1(t)\,dt = 0.$$

Suppose $\epsilon > 0$ and choose $A_0 > t_0$ such that

$$\left| \int_A^{A+\lambda} f_1(t)\,dt \right| < \lambda\epsilon/2, \qquad A > A_0.$$

Corresponding to each $B > A_0$, let $\varphi(B)$ and $N(B)$ be numbers such that $A_0 \leqq \varphi(B) < A_0 + \lambda$, $N(B)$ is an integer, and

$$B = \varphi(B) + N(B)\lambda.$$

Then, when $B > A_0$,

$$\int_{t_0}^{B} f_1(t)\,dt = \int_{t_0}^{\varphi(B)} f_1(t)\,dt + \sum_{n=1}^{N(B)} \int_{\varphi(B)+(n-1)\lambda}^{\varphi(B)+n\lambda} f_1(t)\,dt$$

so that

$$\left| \frac{1}{B} \int_{t_0}^{B} f_1(t)\,dt \right| \leqq \frac{M}{B} + \frac{N(B)\lambda\epsilon}{2B}, \tag{6.24}$$

where M denotes the maximum over the interval $A_0 \leqq u \leqq A_0 + \lambda$ of the continuous function

$$\left| \int_{t_0}^{u} f_1(t)\, dt \right|.$$

Hence the left member of (6.24) converges to 0 as $B \to \infty$; and, using the fact that $f_1(t) = f(t) - L$, we obtain

$$\lim_{B \to \infty} \frac{1}{B} \int_{t_0}^{B} f(t)\, dt = L.$$

Since the limit in (6.21) is λL, we obtain (6.23). The case in which $\lambda \leqq 0$ now follows easily and Theorem 6.2 is proved.

Certain functions of the form

$$f(t) = L + t^{\alpha} \sin t^{\beta}$$

show that the hypothesis that (6.01) holds for each λ does not imply that $f(t)$ is bounded. Other examples of functions $f(t)$ for which (6.01) holds for each λ have the form $f(t) = n\sigma_n$ when n is a positive integer and $n \leqq t < n + n^{-2}$ and $f(t) = 0$ otherwise, the numbers $\sigma_1, \sigma_2, \cdots$ forming a real bounded sequence. It is easy to determine the sequence σ_n in such a way that the functions $f(t)$ and

$$\int_0^B f(t)\, dt$$

each have, as their arguments become infinite, inferior limit $-\infty$ and superior limit $+\infty$.

When $f(t)$ is not assumed bounded, the hypothesis that (6.01) holds for two values λ_1 and λ_2 of λ for which λ_1/λ_2 is irrational does not imply that (6.01) holds for each real λ. This is established by the following example. Let $n_1, n_2, n_3, \cdots$ be an increasing sequence of positive integers for which $n_{p+1}/n_p \to \infty$ as $p \to \infty$. Suppose, to simplify typography,

$$d_p = 2^{-n_p}, \qquad a_p = 4^{n_p}, \qquad b_p = 4^{n_p+1}.$$

Suppose $f(t) = 0$ except when t is a point of one of the intervals $a_p < t < b_p$. For each $p = 1, 2, 3, \cdots$, let $f(x)$ be defined over the interval $a_p < t < b_p$ as follows. For each integer k for which $1 \leqq k \leqq (b_p - a_p)/2d_p$, suppose

$$f(x) = (-1)^k 2k/(b_p - a_p), \qquad a_p + (k-1)d_p < x \leqq a_p + kd_p;$$

and let $f(x)$ be defined over the remaining half of the interval by the formula

$$f(b_p - x) = -f(a_p + x), \qquad 0 < x < \tfrac{1}{2}(b_p - a_p).$$

For each $j = 1, 2, 3, \cdots$, the limit in (6.01) exists and is 0 when $\lambda = 2^{-j}$; an examination of the graph of $f(t)$ indicates the manner in which proof proceeds. The limit in (6.01) also exists and is 0 when λ is the irrational number μ defined by

$$\mu = \sum_{p=1}^{\infty} 2^{1-n_p};$$

proof of this depends upon the fact that μ/d_p is only a little greater than an even integer when p is large, that is,

$$\lim_{p\to\infty} (\mu/d_p - 2[\mu/2d_p]) = 0,$$

where $[x]$ denotes the greatest integer in x. That the conclusion of Theorem 3.1 fails to hold for this function is an obvious consequence of the fact that, for each $p = 1, 2, \cdots$, there is an interval of length d_p over which the integral of $f(t)$ is 1 and another of the same length over which the integral is -1. Hence Theorems 2.1 and 3.1 imply that the set of λ for which the limit in (6.01) exists must have measure 0. It can in fact be proved that, for this example, the limit in (6.01) exists for a given λ if and only if λ can be represented in the form

$$\lambda = \lambda_0 + \sum_{j=1}^{\infty} \theta_j 2^{-j},$$

where λ_0 is an integer, each θ_j is 0 or 1, and a sequence m_p of integers exists such that $m_p \to \infty$ and

$$\theta_j = \theta_{n_p}, \qquad n_p \leqq j \leqq n_p + m_p,$$

for each $p = 1, 2, 3, \cdots$.

CORNELL UNIVERSITY.

CLASSIFICATION OF SOLUTIONS AND OF PAIRS OF SOLUTIONS OF $y''' + 2py' + p'y = 0$ BY MEANS OF INITIAL CONDITIONS

By Joseph J. Eachus

The following facts concerning real solutions of the real differential equation $y''' + 2py' + p'y = 0$ are either explicitly stated in or are readily deducible from the work of G. D. Birkhoff [Annals of Mathematics, (2), vol. 12(1911), pp. 103–127]. There are non-vanishing solutions. If a solution has a double zero,[1] it has no simple zeros. If each of two linearly independent solutions has double zeros, their zeros interlace along the x-axis. If one of two solutions has double zeros while the second has simple zeros, either each zero of the first coincides with a zero of the second, there being exactly one zero of the second between successive coincidences, or the solutions have no zeros in common, there being exactly two zeros of the second between successive zeros of the first. If each of two linearly independent solutions has simple zeros, either their zeros interlace, or they do so in pairs, or alternate zeros of the first coincide with alternate zeros of the second.

It is the purpose of this paper to distinguish between the three types of solutions, and between the various situations with regard to two solutions, by means of numbers determined by the values of the solutions and their first two derivatives at any given point. In order to achieve this end, it is necessary to demonstrate in a manner different from that of Birkhoff that the above statements are true.

The author gratefully acknowledges suggestions made by R. D. Carmichael, leading to the preparation of this paper.

Let $p \equiv p(x)$, on the interval (a, b), be a real continuous function of the real variable x, with a continuous derivative. Consider

$$y''' + 2py' + p'y = 0. \tag{1}$$

We postulate that some solution of this equation has at least two double zeros on (a, b). It will become apparent in the course of discussion that this is equivalent to demanding that some solution have at least three zeros on (a, b)

Let y_i and y_j be any two solutions of (1). Then

$$y_i(y_j''' + 2py_j' + p'y_j) + y_j(y_i''' + 2py_i' + p'y_i)$$
$$= y_iy_j''' + y_i'''y_j + 2py_iy_j' + 2py_i'y_j + 2p'y_iy_j = 0.$$

By integration we obtain

$$y_iy_j'' + y_i''y_j - y_i'y_j' + 2py_iy_j = C_{ij}, \tag{2}$$

Received June 9, 1941; in revised form, December 30, 1941. Presented to the American Mathematical Society, December, 1940.

[1] The terms "double zero" and "simple zero" are used in the sense that x_1 is a double zero of y_1 if $y_1(x_1) = y_1'(x_1) = 0$, $y_1''(x_1) \neq 0$, and x_1 is a simple zero of y_2 if $y_2(x_1) = 0$, $y_2'(x_1) \neq 0$.

where C_{ij} is the constant of integration. If $y_i \equiv y_j$, then

$$2y_iy_i'' - y_i'^2 + 2py_i^2 = C_{ii} \equiv C_i. \tag{3}$$

Now let $y_k = y_i + \lambda y_j$, where λ is any constant, and let C_j and C_k be the analogues of C_i for y_j and y_k. Then

$$\begin{aligned} C_k &= 2(y_i + \lambda y_j)(y_i'' + \lambda y_j'') - (y_i' + \lambda y_j')^2 + 2p(y_i + \lambda y_j)^2 \\ &= 2y_iy_i'' - y_i'^2 + 2py_i^2 + 2\lambda(y_iy_j'' + y_i''y_j - y_i'y_j' + 2py_iy_j) \\ &\qquad + \lambda^2(2y_jy_j'' - y_j'^2 + 2py_j^2) = C_i + 2\lambda C_{ij} + \lambda^2 C_j. \end{aligned}$$

Thus we have

$$C_k = C_i + 2\lambda C_{ij} + \lambda^2 C_j, \tag{4}$$

if $y_k = y_i + \lambda y_j$.

Let us define one more constant, to be associated with a pair of solutions:[2]

$$D_{ij} \equiv C_{ij} - C_iC_j. \tag{5}$$

From (3) it is evident that if y_1 has a double zero, $C_1 = 0$, and if y_1 has a simple zero, $C_1 < 0$, and therefore no solution has both double and simple zeros. It is likewise evident that y_1 has no zeros if $C_1 > 0$. Under the postulate that some solution has two or more double zeros, it is also true that $C_1 > 0$ if y_1 has no zeros—and in fact if y_1 has no zeros on an interval whose end points are zeros of y_2, y_2 being the postulated solution with two double zeros. For, let x_1 and x_2 be successive (double) zeros of y_2. The function y_2/y_1 and its derivative are continuous over (x_1, x_2) and the function vanishes at the end points of the interval. There must then be a point x_0, $x_1 < x_0 < x_2$, where the derivative vanishes, that is,

$$y_2(x_0)y_1'(x_0) - y_1(x_0)y_2'(x_0) = 0.$$

Write

$$y_3 \equiv \frac{y_2(x_0)y_1 - y_1(x_0)y_2}{y_2(x_0)}.$$

Then

$$y_3(x_0) = y_3'(x_0) = 0, \quad \text{and}$$

$$y_1 = y_3 + \frac{y_1(x_0)}{y_2(x_0)}\, y_2, \tag{6}$$

with $C_2 = 0$, $C_3 = 0$. It follows from (4) that

$$C_1 = \frac{2y_1(x_0)}{y_2(x_0)}\, C_{23}.$$

[2] If y_i and y_j are solutions of (1), then so also is $Y \equiv y_iy_j' - y_i'y_j$ (see Birkhoff). It may be noted that $-D_{ij}$ is the analogue of C_i for Y.

But from (2),

$$C_{23} = y_2(x_1)y_3''(x_1) + y_2''(x_1)y_3(x_1) - y_2'(x_1)y_3'(x_1) + 2p(x_1)y_2(x_1)y_3(x_1)$$
$$= y_2''(x_1)y_3(x_1),$$

since $y_2(x_1) = y_2'(x_1) = 0$.

Also from (6), $y_3(x_1) = y_1(x_1)$ and hence

$$C_1 = \frac{2y_1(x_0)y_1(x_1)y_2''(x_1)}{y_2(x_0)}. \tag{7}$$

Since y_1 is non-vanishing over (x_1, x_2), $y_1(x_0)$ and $y_1(x_1)$ have the same sign. As $y_2(x_0)$ and $y_2''(x_1)$ must also have the same sign, it follows that $C_1 > 0$.

Since the conditions $C_1 = 0$, $C_1 < 0$, $C_1 > 0$, are exhaustive and mutually exclusive, as are the conditions y_1 has double zeros, y_1 has simple zeros, y_1 has no zeros, we may state

THEOREM I. *If y_1 is a solution of* (1), *and there is some solution of* (1) *having at least two double zeros, then*

(A) $C_1 = 0$, *if and only if y_1 has double zeros;*

(B) $C_1 < 0$, *if and only if y_1 has simple zeros;*

(C) $C_1 > 0$, *if and only if y_1 has no zeros.*

Consider now two solutions, each having double zeros. Neither can fail to vanish between successive zeros of the other, and hence their zeros alternate along the x-axis. If we refer to the zeros of a solution having double zeros as a "set", any two sets interlace.

Suppose now that y_1 is some solution having simple zeros, and we look for solutions y_i such that $C_i = 0$, $C_{1i} = 0$. This is most easily attacked by examining (2) and (3) at some point x_1 which is a zero of y_1. Here we require that

$$y_i(x_1)y_1''(x_1) - y_i'(x_1)y_1'(x_1) = 0 \tag{8}$$

and

$$2y_i(x_1)y_i''(x_1) - (y_i'(x_1))^2 + 2p(x_1)(y_i(x_1))^2 = 0. \tag{8'}$$

There are two linearly independent solutions of (1) which satisfy (8) and (8'), one defined by

$$y_i(x_1) = y_i'(x_1) = 0, \qquad y_i''(x_1) = 1,$$

the other by

$$y_i(x_1) = 1, \qquad y_i'(x_1) = \frac{y_1''(x_1)}{y_1'(x_1)}, \qquad y_i''(x_1) = \frac{1}{2}\left(\frac{y_1''(x_1)}{y_1'(x_1)}\right)^2 - p(x_1).$$

Designate these by y_2 and y_3 respectively. By considering (8) and (8') as an algebraic system, it is easily seen that every solution which satisfies (8) and (8') is linearly dependent on y_2 or on y_3. Since $C_2 = 0$, $C_3 = 0$, $C_{12} = 0$, $C_{13} = 0$,

we see from (2) that every (double) zero of y_2 and of y_3 is a zero of y_1. Now every zero of y_1 is either a zero of y_2 or a zero of y_3, for, let x_2 be any zero of y_1 and let y_4 be defined by $y_4(x_2) = y_4'(x_2) = 0$, $y_4''(x_2) = 1$. Then $C_4 = 0$ and $C_{14} = 0$, and y_4 is linearly dependent on y_2 or on y_3, whence either $y_2(x_2) = 0$ or $y_3(x_2) = 0$. Thus, if a solution y_1 has simple zeros, those zeros compose two sets, and

THEOREM II. *If $C_1 < 0$ and $C_2 = 0$, then*

(A) *if and only if $C_{12} = 0$, every zero of y_2 is a zero of y_1, and alternate zeros of y_1 are zeros of y_2;*

(B) *if and only if $C_{12} \neq 0$, successive zeros of y_2 are separated by exactly two zeros of y_1.*

The zeros of y_i compose two sets if $C_i < 0$, and in fact may be any two sets. For, let x_1 be a point of one set and x_2 a point of another. Define y_1, y_2 and y_3 by

$$y_1(x_1) = 0, \qquad y_1'(x_1) = 1, \qquad y_1''(x_1) = 0,$$

$$y_2(x_1) = y_2'(x_1) = 0, \qquad y_2''(x_1) = 1, \qquad y_3 \equiv y_2(x_2)y_1 - y_1(x_2)y_2.$$

Then $y_3(x_1) = y_3(x_2) = 0$, and $y_3 \not\equiv 0$. If $y_4(x_1) = y_4(x_2) = 0$, y_4 is readily seen to be linearly dependent on y_3.

Since the zeros of any two sets interlace, there are only three possible configurations of the zeros of y_1 and y_2 if $C_1 < 0$, $C_2 < 0$, and y_1 and y_2 are linearly independent. Either the zeros of y_1 and those of y_2 separate one another, or they do so in pairs, or alternate zeros of y_1 coincide with alternate zeros of y_2.

Now let $y_3 = y_1 + \lambda y_2$. The three cases may be distinguished by the number of values of λ for which C_3 is zero, $C_3 = C_1 + 2\lambda C_{12} + \lambda^2 C_2$ (see (4)). In the first case—separation of the zeros singly—there are evidently no such values of λ. If there were, y_3 for that λ would be always of one sign, but y_2, and hence y_3, cannot have the same sign at successive zeros of y_1, as there is exactly one intervening (simple) zero of y_2. It follows that in this case, $D_{12} \equiv C_{12}^2 - C_1C_2$ is negative. In the second case—separation of zeros by pairs—there are two values of λ for which $C_3 = 0$. The argument used in establishing Theorem I shows that if one solution does not vanish on an interval and another has zeros at the ends of the interval, there is then a linear combination of the two which has a double zero within the interval. In the present case there must be a λ such that the corresponding y_3 has a double zero between two non-separated zeros of y_1, and a λ such that the corresponding y_3 has a double zero between two non-separated zeros of y_2. These values of λ may not be the same, as the zeros of the respective y_3's are not members of the same set, in view of the interlacing of sets. We have, then, that D_{12} is positive in this case. Finally, if alternate zeros of y_1 coincide with alternate zeros of y_2, D_{12} is zero, and there is one and only one λ for which $C_3 = 0$. For, let x_1 be a coincident zero. From (2) and (3),

$$C_1 = -(y_1'(x_1))^2, \qquad C_2 = -(y_2'(x_1))^2, \qquad C_{12} = -y_1'(x_1)y_2'(x_1)$$

and

$$C_3 = -[(y_1'(x_1))^2 + 2\lambda y_1'(x_1)y_2'(x_1) + \lambda^2(y_2'(x_1))^2]$$
$$= -(y_1'(x_1) + \lambda y_2'(x_1))^2.$$

As $y_2'(x_1) \neq 0$, the statement is established.

Since D_{12} must be positive, negative or zero, we have

THEOREM III. *If $C_1 < 0$ and $C_2 < 0$, then*

(A) *if and only if $D_{12} < 0$, the zeros of y_1 and those of y_2 separate one another;*

(B) *if and only if $D_{12} > 0$, the zeros of y_1 and those of y_2 separate one another by pairs;*

(C) *if and only if $D_{12} = 0$, alternate zeros of y_1 coincide with alternate zeros of y_2.*

PURDUE UNIVERSITY.

STRUCTURE AND CONTINUITY OF MEASURABLE FLOWS

By Warren Ambrose and Shizuo Kakutani

1. **Introduction.** The purpose of this paper is to establish some regularity properties for flows which are assumed to satisfy only measurability conditions. In particular, we are concerned with establishing conditions under which a flow will be isomorphic to a continuous flow or a flow built under a function. For measure spaces which satisfy the two conditions of being properly separable and having a separating sequence of measurable sets (see Definitions 8 and 10),[1] our results can be summed up by saying that every flow different from the identity is isomorphic to a (generalized) flow built under a function and to a continuous flow on a separable metric space with a regular measure; these results are obtained in Theorems 2, 4 and 5. For measure spaces which do not satisfy these conditions, the situation is more complicated, and we refer the reader to the body of the paper. Flows built under a function were first introduced, in a special case, by J. von Neumann[2] and have since been considered by one of the present authors.[3] The significance of this isomorphism of any flow to such a flow is that it gives a kind of normal form for a flow and it makes possible the taking of cross sections and the reduction of various properties of a flow to properties of a single measure preserving transformation on such a cross section.

2. **Definitions and notation.**

Definition 1. A *measure space* $\Omega(\mathfrak{B}, m)$ is a system of a space Ω, a Borel field[4] $\mathfrak{B}$ of subsets M of Ω, and a countably additive measure[5] $m(M)$ defined on $\mathfrak{B}$ and satisfying the following conditions:

(2.1) $\Omega \in \mathfrak{B}$ and $m(\Omega) < +\infty$,

(2.2) there exists an $M \in \mathfrak{B}$ for which $0 < m(M) < m(\Omega)$,

Received June 13, 1941.

[1] All measures usually considered, and in particular Lebesgue measures in Euclidean spaces, satisfy these conditions.

[2] J. von Neumann [4], pp. 636–641.

[3] W. Ambrose [1].

[4] A collection of subsets of a space is called a *field* if it is closed under the operations of finite addition, finite intersection and complementation. It is a *Borel field* if it is a field and is closed under the operations of countable addition and intersection. It is easy to see that for any collection of subsets of a space there exists a smallest field, and also a smallest Borel field, which contains the given collection; these are called respectively the field *determined* by the collection and the Borel field *determined* by the collection.

[5] A *countably additive measure* is a non-negative set function $m(M)$ defined for all sets M in some Borel field $\mathfrak{B}$ and having the property that $m\left(\sum_{n=1}^{\infty} M_n\right) = \sum_{n=1}^{\infty} m(M_n)$ for any sequence $\{M_n\}$ $(n = 1, 2, \cdots)$ of disjoint sets from $\mathfrak{B}$.

(2.3) $m(M)$ is *completed*,[6] i.e., if $M \in \mathfrak{B}$ and $m(M) = 0$, then every subset of M is also in $\mathfrak{B}$.

Throughout the present paper we shall use the symbol Ω, with various appendages (e.g., Ω', $\bar{\Omega}$, Ω^*) for a measure space. The corresponding Borel field and measure, as well as points and sets in the space, will always be accompanied by the same appendages. We shall use the symbols Λ, M and N for sets in a measure space, and ω and v for points in such a space. Sets in $\mathfrak{B}$ will be called *measurable* and any set of measure zero will be called a *null set*. If the symmetric difference[7] $M \ominus N$ of two sets M and N is a null set, then we shall say that M and N are *equivalent* and write $M \sim N$. A real valued function $f(\omega)$ defined on a measure space is *measurable* if the set[8] $[\omega: f(\omega) > \alpha]$ is measurable for any real number α.

DEFINITION 2. A *measure preserving transformation* is a one-to-one mapping T of a measure space $\Omega(\mathfrak{B}, m)$ onto a measure space $\Omega'(\mathfrak{B}', m')$ with the property that $M \in \mathfrak{B}$ if and only if $T(M) \in \mathfrak{B}'$, and further that $m(M) = m'(T(M))$ for any measurable set $M \in \mathfrak{B}$. Usually $\Omega(\mathfrak{B}, m)$ and $\Omega'(\mathfrak{B}', m')$ will be the same space. In this case, a set $M \in \mathfrak{B}$ is called *invariant* under T if $\omega \in M$ implies that both $T(\omega) \in M$ and $T^{-1}(\omega) \in M$. A measure preserving transformation T (of a measure space onto itself) is *ergodic* if there exist no measurable sets invariant under T except null sets and complements of null sets.

DEFINITION 3. A *flow* is a one-parameter group $\{T_t\}$ $(-\infty < t < +\infty$, $T_s(T_t(\omega)) = T_{s+t}(\omega)$ for all $s, t, \omega)$ of measure preserving transformations T_t of a measure space onto itself. If $\{T_t\}$ is a flow on a measure space Ω and if ω is a point of Ω, then we shall usually denote $T_t(\omega)$ by ω_t. For a fixed ω, the set of all ω_t $(-\infty < t < +\infty)$ is called the *trajectory* through ω. A set is called *invariant* under a flow if it is invariant under each member of the flow, or equivalently, if whenever it contains a point ω it contains the entire trajectory through ω. A flow is *ergodic* if there exist no invariant measurable sets except null sets and complements of null sets.

DEFINITION 4. A flow $\{T_t\}$ on a measure space Ω is *measurable* if for any measurable set M in Ω the (ω, t)-set M^* defined by $M^* = [(\omega, t): \omega_t \in M]$ is measurable in the product space Ω^* of Ω with the real t-axis, where the measure (and measurability) on Ω^* is defined multiplicatively in terms of the given measure on Ω and the Lebesgue measure on the t-axis.[9]

DEFINITION 5. Let $\{S_t\}$ and $\{T_t\}$ be two flows on measure spaces Ω' and Ω'' respectively. $\{S_t\}$ and $\{T_t\}$ are *isomorphic* if it is possible to split Ω' and

[6] It is not always assumed that a measure has this property, but it is possible to extend the domain of definition of any countably additive measure (i.e., to enlarge the Borel field $\mathfrak{B}$ on which it is defined) to obtain a completed measure.

[7] The symmetric difference of two sets M and N is defined by $M \ominus N = M + N - M \cdot N = (M - MN) + (N - MN)$.

[8] We use the symbol $[\omega; C]$ for the ω-set on which the condition C is satisfied.

[9] For a definition of measure in product space, see S. Saks [5].

Ω'' into two parts $N', \Omega' - N'$ and $N'', \Omega'' - N''$ respectively, in such a way that
(2.4) N' is a null set invariant under $\{S_t\}$ and N'' is a null set invariant under $\{T_t\}$,
(2.5) there exists a measure preserving transformation of $\Omega' - N'$ onto $\Omega'' - N''$ which carries $\{S_t\}$ into $\{T_t\}$, i.e., such that if $\omega' \in \Omega' - N'$ corresponds to $\omega'' \in \Omega'' - N''$ then $S_t(\omega')$ corresponds to $T_t(\omega'')$ for all t.

DEFINITION 6. Let $\Omega(\mathfrak{B}, m)$ be a measure space, and let S be a measure preserving transformation of Ω onto itself. Let $f(\omega)$ be a positive real valued function defined on Ω which is measurable and integrable on Ω. We assume that $\sum_{n=0}^{\infty} f(S^n(\omega)) = +\infty$ and $\sum_{n=1}^{\infty} f(S^{-n}(\omega)) = +\infty$ for any $\omega \in \Omega$. (This condition is surely satisfied if there exists a constant $c > 0$ such that $f(\omega) > c$ for all $\omega \in \Omega$.) Consider the product space of Ω with the real u-axis, with the measure $\bar{m}$ defined on it multiplicatively in terms of m-measure on Ω and the Lebesgue measure on the u-axis. Let $\bar{\Omega}$ be the portion of this product space[10] under the graph of $f(\omega)$, i.e., the set of all points $\bar{\omega} = (\omega, u)$ for which $0 \leqq u < f(\omega)$, and let $\bar{\mathfrak{B}}$ be the collection of all $\bar{m}$-measurable subsets of $\bar{\Omega}$. Then $\bar{\Omega}(\bar{\mathfrak{B}}, \bar{m})$ is a measure space. Define the flow $\{T_t\}$ on $\bar{\Omega}$ by

$$
\begin{aligned}
T_t(\omega, u) &= (\omega, u + t), \quad \text{if} \quad -u \leqq t < -u + f(\omega), \\
&= (S^n(\omega), u + t - f(\omega) - \cdots - f(S^{n-1}(\omega))), \\
&\qquad \text{if} \quad -u + \sum_{k=0}^{n-1} f(S^k(\omega)) \leqq t < -u + \sum_{k=0}^{n} f(S^k(\omega)), \\
&\qquad\qquad n = 1, 2, \cdots, \qquad (2.6) \\
&= (S^{-n}(\omega), u + t + f(S^{-1}(\omega)) + \cdots + f(S^{-n}(\omega))), \\
&\qquad \text{if} \quad -u - \sum_{k=1}^{n} f(S^{-k}(\omega)) \leqq t < -u - \sum_{k=1}^{n-1} f(S^{-k}(\omega)), \\
&\qquad\qquad n = 1, 2, \cdots.
\end{aligned}
$$

We call $\{T_t\}$ the *flow built under the function* $f(\omega)$ *on the measure preserving transformation* S, or simply, a *flow built under a function*. Ω is called the *base space*, S is a *base transformation* and $f(\omega)$ is called a *ceiling function*.

This definition was given in [1]. In the following discussions, it is necessary to consider a general case in which the base space Ω has an infinite measure. In this case we call $\{T_t\}$ a *generalized flow built under a function*. It is, however, to be noticed that we always assume that $\bar{\Omega}$ has a finite measure: $\bar{m}(\bar{\Omega}) = \int_{\Omega} f(\omega)\, dm(\omega) < +\infty$. Hence it is impossible, in this case, that there exist a constant $c > 0$ such that $f(\omega) > c$ for all $\omega \in \Omega$.[11]

[10] Since we assumed that $f(\omega)$ is integrable on Ω, $\bar{\Omega}$ has also a finite measure: $\bar{m}(\bar{\Omega}) = \int_{\Omega} f(\omega)\, dm(\omega) < \infty$.

[11] It is assumed that Ω is a sum of a countable number of subsets of a finite measure. More precisely, we shall consider only the following cases: Let $\{\{T_t^{(n)}\}\}$ $(n = 1, 2, \cdots)$ be a

As in [1] we shall have to consider the functions $F(\tilde{\omega})$ and $G(\tilde{\omega})$ associated with a (generalized) flow built under a function, and defined by

$$F(\tilde{\omega}) = F(\omega, u) = f(\omega), \tag{2.7}$$

$$G(\tilde{\omega}) = G(\omega, u) = u. \tag{2.8}$$

These are both $\bar{m}$-measurable functions defined on $\bar{\Omega}$.

3. **Properties of measurable flows.** We begin with two lemmas concerning measurable flows.

LEMMA 1. *Let $\{T_t\}$ be a measurable flow on a measure space $\Omega(\mathfrak{B}, m)$. Then*
(3.1) *for any measurable function $\varphi(\omega)$ defined on Ω, there exists an invariant null set $N \in \mathfrak{B}$ such that the t-function $\varphi(\omega_t)$ is Lebesgue measurable for any $\omega \in \Omega - N$;*
(3.2) *for any null set $M \in \mathfrak{B}$ there exists an invariant null set $N \in \mathfrak{B}$ such that the t-set $[t: \omega_t \in M]$ is of Lebesgue measure zero for any $\omega \in \Omega - N$.*

Proof. These are immediate consequences of the definition of a measurable flow and Fubini's theorem.

LEMMA 2. *Let $\{T_t\}$ be a measurable flow on a measure space $\Omega(\mathfrak{B}, m)$, and let $M \in \mathfrak{B}$ be a measurable set such that $T_t(M) \sim M$ for every t. Then there exists an invariant measurable set $N \in \mathfrak{B}$ such that $N \sim M$.*

Proof. See E. Hopf [3], p. 27.

DEFINITION 7. A flow $\{T_t\}$ on a measure space $\Omega(\mathfrak{B}, m)$ is *proper* if every measurable set of positive measure contains a measurable set $M \in \mathfrak{B}$ such that $m((\Omega - M) \cdot T_{t_0}(M)) > 0$ for some t_0; it is *completely improper* if $M \sim T_t(M)$ for any measurable set $M \in \mathfrak{B}$ and for any t. It is a consequence of Lemma 2 that a flow is completely improper if and only if every measurable set is equivalent to an invariant set.

THEOREM 1. *Let $\{T_t\}$ be a measurable flow on a measure space Ω. Then $\Omega = \Omega_1 + \Omega_2$, where Ω_1 and Ω_2 are disjoint invariant measurable sets, and $\{T_t\}$ is completely improper on Ω_1 and proper on Ω_2.*

---

sequence of flows built under a function, each defined on a measure space $\bar{\Omega}_n(\bar{\mathfrak{B}}, \bar{m})$. We denote the base space of $\bar{\Omega}_n$ by $\Omega_n(\mathfrak{B}, m)$. The base transformation and the ceiling function on Ω_n are denoted by S and $f(\omega)$ respectively (without suffix n). We assume that $\sum_{n=1}^{\infty} \bar{m}(\bar{\Omega}_n) < +\infty$, but we do not assume that $\sum_{n=1}^{\infty} m(\Omega_n) < +\infty$. Let us put $\bar{\Omega} = \sum_{n=1}^{\infty} \bar{\Omega}_n$ and $\Omega = \sum_{n=1}^{\infty} \Omega_n$ (in each case, summands are assumed to be disjoint). $\bar{\Omega}(\bar{\mathfrak{B}}, \bar{m})$ is a measure space, while this is not always true for $\Omega(\mathfrak{B}, m)$ since $m(\Omega) = \sum_{n=1}^{\infty} m(\Omega_n)$ may be infinite. Then, by putting together these flows $\{\{T_t^{(n)}\}\}$ $(n = 1, 2, \cdots)$, we shall have a flow $\{T_t\}$ defined on $\bar{\Omega}$; indeed, $\{T_t\}$ is defined by $T_t(\bar{\omega}) = T_t^{(n)}(\bar{\omega})$ if $\bar{\omega} \in \bar{\Omega}_n$. $\{T_t\}$ is a generalized flow built under the function $f(\omega)$ on the transformation S. This differs from the ordinary flow built under a function only in that we do not assume the finiteness of $m(\Omega) = \sum_{n=1}^{\infty} m(\Omega_n)$.

Proof. If $\{T_t\}$ is not proper on Ω, then there exists a measurable set M of positive measure such that every measurable subset of M is equivalent to an invariant set. By Lemma 2, we may assume that M itself is invariant. Let $\mathfrak{A}$ be the collection of all such sets M, and let α be the least upper bound of $m(M)$ for all $M \in \mathfrak{A}$. Then there exists a sequence $\{M_n\}$ $(n = 1, 2, \cdots)$ of measurable sets from $\mathfrak{A}$ for which $m(M_n) \to \alpha$. Putting $\Omega_1 = \sum_{n=1}^{\infty} M_n$ and $\Omega_2 = \Omega - \Omega_1$, it is clear that we have the required decomposition.

This theorem is useful because it allows us to consider separately the completely improper flows and the proper flows, and to forget about the intermediate case of flows which are neither proper nor completely improper. We shall consider these two cases in §§4 and 5 respectively.

4. **Structure of completely improper flows.** Obviously, the identity flow (defined by $\omega_t = \omega$ for all ω and t) is completely improper, and the question arises whether there are any other completely improper flows. That there are others is shown by the following example. Consider a flow built under some function on the identity transformation (the particular function and the particular space on which the identity transformation is defined do not matter); but, instead of taking for our measurable sets the usual collection of all $\bar{m}$-measurable sets, we take a subcollection of those $\bar{m}$-measurable sets which are equivalent to a set depending on ω alone.[12] Then it is clear that this is a completely improper flow, while there is no point which is invariant under the flow.

In this example of a completely improper flow, we first took a proper flow (it will be shown in Theorem 4 that every flow built under a function is proper) and then decreased the collection of measurable sets to obtain a completely improper flow. This fact leads to the suspicion that when a flow is improper it may not be through any fault of the flow itself but rather may be due to some deficiency in the collection of measurable sets of the measure space on which it is defined. Theorem 3 below confirms this suspicion to a certain extent by determining the general form of completely improper flows, and Theorem 2 will show that in all measure spaces usually considered the only completely improper measurable flow is the identity.

DEFINITION 8. A countable collection $\{M_n\}$ $(n = 1, 2, \cdots)$ of subsets of Ω is called a *separating sequence*, if for every pair of points ω and v of Ω $(\omega \neq v)$ there exists an M_n which contains one but not both of ω and v.

THEOREM 2. *Let $\{T_t\}$ be a completely improper measurable flow defined on a measure space $\Omega(\mathfrak{B}, m)$. If Ω has a separating sequence of measurable sets, then there exists an invariant null set $N \in \mathfrak{B}$ such that $\{T_t\}$ is the identity on $\Omega - N$.*

Proof. Let $\{M_n\}$ $(n = 1, 2, \cdots)$ be a separating sequence of measurable sets in Ω. Since $\{T_t\}$ is completely improper, there exists a sequence of in-

[12] A set $\bar{M} \subset \bar{\Omega}$ depends on ω alone if $(\omega, u) \in \bar{M}$ implies $(\omega, v) \in \bar{M}$ for all v with $0 \leq v < f(\omega)$; such a set $\bar{M}$ is necessarily of the form $(M \times (-\infty, +\infty)) \cdot \bar{\Omega}$; we say that $\bar{M}$ is determined by M.

variant measurable sets $\{\Lambda_n\}$ $(n = 1, 2, \cdots)$ such that $\Lambda_n \sim M_n$ $(n = 1, 2, \cdots)$. We define a null set M by $M = \sum_{n=1}^{\infty} (M_n \ominus \Lambda_n) = \sum_{n=1}^{\infty} (M_n + \Lambda_n - M_n \cdot \Lambda_n)$. Then, by Lemma 1, there exists an invariant null set $N \in \mathfrak{B}$ such that for $\omega \in \Omega - N$ the t-set $M(\omega) = [t: \omega_t \in M]$ is of Lebesgue measure zero. We shall prove that this N is a required set, i.e., that $\omega_t = \omega$ for any $\omega \in \Omega - N$ and for any t. First we prove that, for $\omega \in \Omega - N$, $t_1 \in' M(\omega)$ and $t_2 \in' M(\omega)$ imply $\omega_{t_1} = \omega_{t_2}$. Indeed, if we have $\omega_{t_1} \neq \omega_{t_2}$ for such t_1 and t_2, then there exists an M_n which contains one but not both of ω_{t_1} and ω_{t_2}. Since ω_{t_1} and ω_{t_2} both do not belong to $M_n \ominus \Lambda_n$, the same thing must be true of Λ_n. This is, however, a contradiction to the fact that Λ_n is invariant under $\{T_t\}$. Thus we have proved that, for every $\omega \in \Omega - N$, $t_1 \in' M(\omega)$ and $t_2 \in' M(\omega)$ imply $\omega_{t_1} = \omega_{t_2}$. The remainder of the proof of Theorem 2 clearly follows from the following lemma.

LEMMA 3. *Let $\{T_t\}$ be a flow on a measure space Ω. If, for every fixed $\omega \in \Omega$, the t-set E defined by $E = [t: \omega_t = \omega]$ is Lebesgue measurable and of positive measure, then $\omega_t = \omega$ for all t.*

Proof. The group property of the flow clearly implies that E is an additive group of real numbers. The lemma then follows from the known theorem that the only additive group of real numbers which is Lebesgue measurable and of positive measure is the whole real line.

DEFINITION 9. Consider a measure space $\Omega(\mathfrak{B}, m)$ and a real valued non-negative function $f(\omega)$ defined on it. We do not assume that $f(\omega)$ is measurable. Moreover, $f(\omega)$ may be equal to 0 or $+\infty$. For every point $\omega \in \Omega$ consider the set $\bar{\Omega}(\omega)$ of real numbers defined as follows: if $f(\omega) = 0$, then $\bar{\Omega}(\omega)$ consists only of a single point 0; if $0 < f(\omega) < +\infty$, then $\bar{\Omega}(\omega)$ is a semi-open interval: $0 \leqq u < f(\omega)$; if $f(\omega) = +\infty$, then $\bar{\Omega}(\omega)$ is an infinite interval: $-\infty < u < +\infty$. Let $\bar{\Omega}$ be the set of all pairs (ω, u), where $\omega \in \Omega$ and $u \in \bar{\Omega}(\omega)$.

We shall make $\bar{\Omega}$ into a measure space. Let $\bar{\mathfrak{N}}_0$ be the collection of all sets $\bar{N} \subset \bar{\Omega}$ for which there exists a null set $N \in \mathfrak{B}$ $(N \subset \Omega)$ such that $\omega \in \Omega - N$ implies $m_0(\bar{N}(\omega)) = 0$, where $\bar{N}(\omega)$ is the set of all $u \in \bar{\Omega}(\omega)$ such that $\bar{\omega} = (\omega, u) \in \bar{N}$ and m_0 is a measure defined on $\bar{\Omega}(\omega)$ in the following way: if $f(\omega) = 0$, then $m_0(\{0\}) = 1$ ($\{0\}$ is a set consisting of 0 alone); if $0 < f(\omega) \leqq +\infty$, then m_0 is the ordinary Lebesgue measure. Clearly $\bar{\mathfrak{N}}_0$ has the following properties:

$$\bar{N}_n \in \bar{\mathfrak{N}}_0 \ (n = 1, 2, \cdots) \quad \text{imply} \quad \sum_{n=1}^{\infty} \bar{N}_n \in \bar{\mathfrak{N}}_0, \tag{4.1}$$

$$\bar{N}_0 \in \bar{\mathfrak{N}}_0, \quad \bar{N} \subset \bar{N}_0 \quad \text{imply} \quad \bar{N} \in \bar{\mathfrak{N}}_0. \tag{4.2}$$

Now let $\bar{\mathfrak{N}}$ be an arbitrary subcollection of $\bar{\mathfrak{N}}_0$ which has the same properties (4.1) and (4.2). Let $\bar{\mathfrak{B}}_0$ be the collection of all sets $\bar{M} \subset \bar{\Omega}$ which depend on ω alone and are determined by a set $M \in \mathfrak{B}$ $(M \subset \Omega)$. Then the collection $\bar{\mathfrak{B}}$ is defined as follows: a set $\bar{\Lambda} \subset \bar{\Omega}$ belongs to $\bar{\mathfrak{B}}$ if and only if there exists a set $\bar{M} \in \bar{\mathfrak{B}}_0$ such that the symmetric difference $\bar{\Lambda} \ominus \bar{M}$ belongs to $\bar{\mathfrak{N}}$. $\bar{\mathfrak{B}}$ is clearly a

Borel field and if we put $\bar{m}(\bar{\Lambda}) = m(M)$, where $M \in \mathfrak{B}$ is the set in Ω which determined $\bar{M} \in \overline{\mathfrak{B}}_0$, then $\bar{m}(\bar{\Lambda})$ is a countably additive measure which is defined and completed on $\overline{\mathfrak{B}}$. We define a flow $\{T_t\}$ on the measure space $\bar{\Omega}(\overline{\mathfrak{B}}, \bar{m})$ as follows: if $f(\omega) = 0$, then $T_t(\omega, 0) = (\omega, 0)$ for all t (i.e., $\{T_t\}$ is the identity); if $0 < f(\omega) < +\infty$, then $T_t(\omega, u) = (\omega, u + t)$, where the second coordinate is to be taken modulo $f(\omega)$ (i.e., $\{T_t\}$ is isomorphic to a rotation of a circumference); if $f(\omega) = +\infty$, then $T_t(\omega, u) = (\omega, u + t)$ for all t (i.e., $\{T_t\}$ is isomorphic to a translation on an infinite line). It is clear that the flow $\{T_t\}$ thus defined is a completely improper flow. Such a flow is called a *singular flow*.

THEOREM 3. *Every completely improper measurable flow is isomorphic to a singular flow.*

Proof. Let $\{S_t\}$ be a completely improper measurable flow defined on a measure space $\Omega^*(\mathfrak{B}^*, m^*)$. From each trajectory of $\{S_t\}$ pick up a single point ω^*, and let Ω be the set of all such points. When we consider ω^* as a point of Ω, we denote it by ω. We shall define a measure on Ω. A collection $\mathfrak{B}$ of measurable sets in Ω is defined as follows: a subset M of Ω belongs to $\mathfrak{B}$ if and only if the corresponding set $M^* = \sum_{\omega \in M} \Omega^*(\omega)$ belongs to $\mathfrak{B}^*$, where we denote by $\Omega^*(\omega)$ the trajectory through a point ω^* which corresponds to ω. $\mathfrak{B}$ is clearly a Borel field, and if we put $m(M) = m^*(M^*)$ for every $M \in \mathfrak{B}$, then $m(M)$ is a countably additive measure which is defined and completed on $\mathfrak{B}$. Thus we have defined a measure space $\Omega(\mathfrak{B}, m)$.

Now we shall define a function $f(\omega)$ on Ω as follows: if $\Omega^*(\omega)$ consists of ω^* alone (i.e., if ω^* is invariant under $\{S_t\}$), then $f(\omega) = 0$; if $\{S_t\}$ is periodic on $\Omega^*(\omega)$, then $f(\omega) =$ period of $\{S_t\}$ on $\Omega^*(\omega)$; if $\{S_t\}$ has no period on $\Omega^*(\omega)$ (i.e., if $S_{t_1}(\omega^*) \neq S_{t_2}(\omega^*)$ for any t_1 and t_2 $(t_1 \neq t_2)$), then $f(\omega) = +\infty$. We construct the measure space $\bar{\Omega}(\overline{\mathfrak{B}}, \bar{m})$ in terms of the measure space $\Omega(\mathfrak{B}, m)$ and the function $f(\omega)$ thus defined as in Definition 9. (Here it must be noticed that there was a certain arbitrariness in the choice of a subcollection $\overline{\mathfrak{N}}$ of $\overline{\mathfrak{N}}_0$.) We shall now show that, if we take a suitable subcollection $\overline{\mathfrak{N}}$ of $\overline{\mathfrak{N}}_0$, then the singular flow $\{T_t\}$ thus obtained is isomorphic to the given flow $\{S_t\}$. In order to prove this, consider the correspondence $S_t(\omega^*) \leftrightarrow (\omega, t)$, where $w^* \in \Omega^*$ is a point chosen from each trajectory which corresponds to a point $\omega \in \Omega$. It is clear that the flows $\{S_t\}$ and $\{T_t\}$ are carried over into each other by this correspondence, and it only remains to show that the Borel fields $\mathfrak{B}^*$ and $\overline{\mathfrak{B}}$ are also carried over into each other by this correspondence. This is, however, a consequence of the following two facts: (1) by the definition of a completely improper flow, every measurable set $\Lambda^* \in \mathfrak{B}^*$ is equivalent to an invariant set $M^* \in \mathfrak{B}^*$ and the latter corresponds to a set $\bar{M} \in \overline{\mathfrak{B}}_0$ which depends on ω alone; and (2) by Lemma 1, every set $N^* \in \mathfrak{B}^*$ of measure zero corresponds to a set $\bar{N} \in \overline{\mathfrak{N}}_0$, where $\overline{\mathfrak{N}}_0$ is a collection defined in Definition 9, and the collection of all such sets $\bar{N}$ (which correspond to some null set $N^* \in \mathfrak{B}^*$) has the properties (4.1) and (4.2). Hence we have only to take $\overline{\mathfrak{N}}$ as the collection of all such sets $\bar{N}$. This proves Theorem 3.

5. **Structure of proper flows; the fundamental representation theorem.** In this section we prove that a measurable flow is isomorphic to a generalized flow built under a function if and only if it is proper. This theorem was proved previously for ergodic flows by one of us [1]. In order to prove it for non-ergodic case, we need Lemma 4 below, which is essentially due to Poincaré.[13] We remark that the proof for the ergodic case was obtained by using a certain recurrence property that follows from ergodicity, whereas this lemma gives us a usable recurrence without any such assumption as ergodicity.

LEMMA 4. *Let $\{T_t\}$ be a flow on a measure space Ω, and let Λ be any measurable set in Ω. Then there exists a subset N of Ω of measure zero such that for any $\omega \in \Lambda - N$ the trajectory through ω intersects Λ infinitely often both as $t \to +\infty$ and as $t \to -\infty$.*

Proof. This lemma follows trivially by applying Hilfssatz 13.3, p. 48 of E. Hopf [3], to any member, other than T_0, of the flow $\{T_t\}$.

THEOREM 4. *A measurable flow is isomorphic to a generalized flow built under a function if and only if it is proper.*[14]

Proof. First we prove that a generalized flow built under a function (and hence any flow isomorphic to a generalized flow built under a function) is proper. Let $\{T_t\}$ be a generalized flow built under a function (we use the notation of Definition 6) and let $\bar{\Lambda}$ be any measurable set of positive measure. Then choose real numbers a and b with $0 < b - a < a$ such that the set $\bar{M}$ defined by

$$\bar{M} = [\bar{\omega}: a < G(\bar{\omega}) < b] \cdot \bar{\Lambda}$$

has a positive measure (that this can be done is a consequence of the fact that the measure $\bar{m}$ on the (ω, u)-space is defined multiplicatively in terms of the measure m of Ω and the Lebesgue measure on the u-axis). Then $\bar{M}$ is a measurable subset of $\bar{\Lambda}$ with the property that $\bar{m}((\bar{\Omega} - \bar{M}) \cdot T_{-a}(\bar{M})) > 0$. This proves that $\{T_t\}$ is proper.

Now we prove that, conversely, every proper measurable flow is isomorphic to a generalized flow built under a function. Let $\{S_t\}$ be a proper measurable flow on a measure space $\Omega^*(\mathfrak{B}^*, m^*)$. We shall find a decomposition of Ω^* into a sum of mutually disjoint invariant measurable sets $\{\Omega_n^*\}$ $(n = 1, 2, \cdots)$ on each of which the flow $\{S_t\}$ is isomorphic to a flow built under a function. The number of these parts may be either finite or countably infinite, and there may also be a remaining part of measure zero. Putting together these flows built under a function, we finally have a generalized flow built under a function which is isomorphic to $\{S_t\}$.

Consider the collection $\mathfrak{A}^*$ of all measurable sets $M^* \in \mathfrak{B}^*$ of the form: $M^* = (\Omega^* - \Lambda^*) \cdot S_t(\Lambda^*)$, where $\Lambda^* \in \mathfrak{B}^*$ and t is any real number. Let $\delta_1 = \delta(\Omega^*)$ be the upper bound of the measures $m^*(M^*)$ for all $M^* \in \mathfrak{A}^*$. Since $\{S_t\}$ is proper, $\delta_1 = \delta(\Omega^*)$ must be positive. We shall now show that there exists an

[13] See E. Hopf [3], p. 48.

[14] Theorem 4 includes Theorem 2 of [4] as a special case since, as is easily seen, every ergodic flow is proper.

invariant measurable set $\Omega_1^* \subset \Omega^*$ such that $m^*(\Omega_1^*) > \frac{1}{6}\delta_1$ and on which the flow $\{S_t\}$ is isomorphic to a flow built under a function.

For this purpose, let $M^* \in \mathfrak{M}^*$ be a measurable set with $m^*(M^*) > \frac{1}{2}\delta_1$, and let Λ^* be a measurable set such that $M^* = (\Omega^* - \Lambda^*) \cdot S_{t_0}(\Lambda^*)$, where t_0 is a suitable real number. We shall denote the characteristic function of Λ^* by $\varphi(\omega^*)$. Then, by Lemma 1, there exists an invariant set N^* of measure zero such that for any $\omega^* \in \Omega^* - N^*$ the t-function $\varphi(\omega_t^*)$ is Lebesgue measurable. For any $\omega^* \in \Omega^* - N^*$ we define a function $\Phi_c(\omega^*)$ by

$$\Phi_c(\omega^*) = \frac{1}{c}\int_0^c \varphi(\omega_t^*)\, dt, \qquad 0 < c \leqq 1, \tag{5.1}$$

and we define the sets Λ_1^* and Λ_2^* by

$$\Lambda_1^* = [\omega^* : \Phi_c(\omega^*) < \tfrac{1}{4}], \qquad \Lambda_2^* = [\omega^* : \Phi_c(\omega^*) > \tfrac{3}{4}]. \tag{5.2}$$

(We shall choose a fixed value of c in a moment.) By a theorem of N. Wiener,[15] we know that $\Phi_c(\omega^*) \to \varphi(\omega^*)$ almost everywhere as $c \to 0$. Hence we can choose a real number c with $0 < c \leqq 1$ in such a way that $m^*(\Lambda_1^* \ominus (\Omega^* - \Lambda^*)) < \frac{1}{3}m^*(M^*)$ and $m^*(\Lambda_2^* \ominus \Lambda^*) < \frac{1}{3}m^*(M^*)$. From these follow easily that $m^*(\Lambda_1^* \cdot M^*) > \frac{2}{3}m^*(M^*)$ and $m^*(S_{t_0}(\Lambda_2^*) \cdot M^*) > \frac{2}{3}m^*(M^*)$, and hence that $m^*(\Lambda_1^* \cdot S_{t_0}(\Lambda_2^*)) > \frac{1}{3}m^*(M^*)$. We now define Ω_1^* to be the set of all points $\omega^* \in \Omega^*$ whose trajectories intersect $\Lambda_1^* \cdot S_{t_0}(\Lambda_2^*)$ infinitely often both as $t \to +\infty$ and as $t \to -\infty$. Obviously Ω_1^* is an invariant set, and it is formally given by

$$\Omega_1^* = \Big(\prod_s \sum_{t>s} S_t(\Lambda_1^* \cdot S_{t_0}(\Lambda_2^*))\Big)\Big(\prod_s \sum_{t<s} S_t(\Lambda_1^* \cdot S_{t_0}(\Lambda_2^*))\Big), \tag{5.3}$$

where s and t in these summations and products run through all real numbers. Using the fact that $\Phi_c(\omega_t^*)$ is a continuous t-function for each $\omega^* \in \Omega^* - N^*$, it is easy to see that the expression on the right hand side of (5.3) is the same whether s and t run through all real numbers or only through rational numbers. Hence Ω_1^* is m^*-measurable. It is clear that $m^*(\Omega_1^*) \geqq m^*(\Lambda_1^* \cdot S_{t_0}(\Lambda_2^*)) > \frac{1}{3}m^*(M^*) > \frac{1}{6}\delta_1$.

Next we shall prove that, on this part Ω_1^*, the flow $\{S_t\}$ is isomorphic to a flow built under a function. To do this we must find a measure space Ω_1, a measure preserving transformation S on Ω_1, and a positive valued function $f(\omega)$ defined on Ω_1 such that the flow built under a function in terms of these is isomorphic to the given flow $\{S_t\}$ on Ω_1^*. First we notice that for every $\omega^* \in \Omega_1^*$ we have (from the definition (5.1) of $\Phi_c(\omega^*)$) that

$$|\Phi_c(\omega_t^*) - \Phi_c(\omega_s^*)| \leqq \frac{2}{c}|t - s|. \tag{5.4}$$

Now we define Ω_1 by

$$\Omega_1 = \Omega_1^*\Big[\omega^* : \Phi_c(\omega^*) = \tfrac{1}{2} \quad \text{and} \quad \Phi_c(\omega_t^*) > \tfrac{1}{2} \quad \text{for all } t \text{ in} \quad 0 < t \leqq \frac{c}{2}\Big]. \tag{5.5}$$

[15] N. Wiener [6], p. 2, Theorem III′.

Since the trajectory of each $\omega^* \in \Omega_1^*$ intersects $\Lambda_1^* \cdot S_{t_0}(\Lambda_2^*)$ infinitely often both as $t \to +\infty$ and as $t \to -\infty$, it will have the same property for each of the sets Λ_1^* and Λ_2^*. Then the fact that $\Phi_c(\omega^*) < \frac{1}{4}$ for $\omega^* \in \Lambda_1^*$ while $\Phi_c(\omega^*) > \frac{3}{4}$ for $\omega^* \in \Lambda_2^*$, together with the continuity of $\Phi_c(\omega_t^*)$ along each trajectory (as a t-function) and the inequality (5.4) imply, as in [1], that each trajectory through a point of Ω_1^* intersects Ω_1 infinitely often both as $t \to +\infty$ and as $t \to -\infty$. Having established this, we can now define S and $f(\omega)$ as follows: for any $\omega \in \Omega_1$ there is a smallest positive number t for which $S_t(\omega) \in \Omega_1$. We define $S(\omega)$ to be $S_t(\omega)$ and $f(\omega)$ to be t. (We notice that $f(\omega)$ is bounded below by a positive constant.)

Now we need to define a measure on Ω_1 for which S is a measure preserving transformation and $f(\omega)$ is a measurable function, and then to show that the flow built under a function $\{T_t\}$ which we obtain from these (as in Definition 6) is isomorphic to the given flow $\{S_t\}$ on Ω_1^*. According to Definition 6, $\{T_t\}$ is defined on a measure space $\bar{\Omega}_1(\overline{\mathfrak{B}}, \bar{m})$ whose points are of the form $\bar{\omega} = (\omega, u)$, $0 \leqq u < f(\omega)$, $\omega \in \Omega_1$. It is clear that the mapping $(\omega, u) \leftrightarrow S_u(\omega)$ gives a one-to-one correspondence between $\bar{\Omega}_1$ and Ω_1^*, and that this correspondence carries $\{T_t\}$ into $\{S_t\}$. Now let $\bar{m}(\bar{M})$ be the measure on $\bar{\Omega}_1$ which is carried over from the m^*-measure on Ω_1^* by this correspondence. Then all we need to show is that this measure $\bar{m}$ is the product measure of a certain measure on Ω_1 with the Lebesgue measure on the u-axis, and where the measure on Ω_1 is such that S is a measure preserving transformation and $f(\omega)$ is a measurable function. To prove this it is sufficient, by Theorem 1 of [1], to show that the functions $F(\bar{\omega})$ and $G(\bar{\omega})$ defined by (2.6) are $\bar{m}$-measurable. Since this can be proved exactly in the same way as in the case of ergodic flows, we omit the details and refer the reader to [1], pp. 734–735.

We have thus proved the existence of an invariant measurable subset Ω_1^* of Ω^*, whose measure satisfies $m^*(\Omega_1^*) > \frac{1}{6}\delta_1 = \frac{1}{6}\delta(\Omega^*)$ and on which the flow $\{S_t\}$ is isomorphic to a flow built under a function.[16] If $\Omega_1^* = \Omega^*$, then our proof is complete; if $m^*(\Omega^* - \Omega_1^*) > 0$, then we consider the flow $\{S_t\}$ on the remaining invariant set $\Omega^* - \Omega_1^*$, on which it is again a proper measurable flow. Then, by the same argument as in above, we can find an invariant measurable subset Ω_2^* of $\Omega^* - \Omega_1^*$ such that $m^*(\Omega_2^*) > \frac{1}{6}\delta_2$ and on which the flow $\{S_t\}$ is isomorphic to a flow built under a function, where $\delta_2 = \delta(\Omega^* - \Omega_1^*)$ is defined for $\Omega^* - \Omega_1^*$ exactly as $\delta_1 = \delta(\Omega^*)$ was defined for Ω^*. Repeating this procedure, if necessary, we obtain at the n-th stage n invariant disjoint measurable subsets $\Omega_1^*, \cdots, \Omega_n^*$ of Ω^*, which are mutually disjoint and on each of which the flow $\{S_t\}$ is isomorphic to a flow built under a function. Moreover, the condition $m^*(\Omega_k^*) > \frac{1}{6}\delta_k = \frac{1}{6}\delta(\Omega^* - \Omega_1^* - \cdots - \Omega_{k-1}^*)$ is satisfied for $k = 1, 2, \cdots, n$. If $m^*(\Omega^* - \Omega_1^* - \cdots - \Omega_n^*) = 0$ for some n, then our proof is complete; if, on the contrary, $m^*(\Omega^* - \Omega_1^* - \cdots - \Omega_n^*) > 0$ for each n, then we have a sequence of invariant measurable subsets $\{\Omega_n^*\}$ $(n = 1, 2, \cdots)$ of Ω^*, which are mutually disjoint

[16] By appealing to transfinite induction we could eliminate the remainder of the proof.

and on each of which the flow $\{S_t\}$ is isomorphic to a flow built under a function. Moreover, the condition $m^*(\Omega_n^*) > \frac{1}{6}\delta_n = \frac{1}{6}\delta(\Omega^* - \Omega_1^* - \cdots - \Omega_{n-1}^*)$ is satisfied for $n = 1, 2, \cdots$. We shall now prove that $m^*(\Omega^* - \sum_{n=1}^{\infty} \Omega_n^*) = 0$. Indeed, if $\Omega_\infty^* = \Omega^* - \sum_{n=1}^{\infty} \Omega_n^*$ is of positive measure, then the flow $\{S_t\}$ is again proper on Ω_∞^* and the corresponding real number $\delta_\infty = \delta(\Omega_\infty^*)$ (defined for Ω_∞^* exactly as $\delta_1 = \delta(\Omega^*)$ was defined for Ω^*) must be positive. This is, however, a contradiction since from $\Omega_\infty^* \subset \Omega^* - \Omega_1^* - \cdots - \Omega_{n-1}^*$ follows $\delta_\infty \leqq \delta_n$ for $n = 1, 2, \cdots$, and hence $n\delta_\infty \leqq \sum_{k=1}^{n} \delta_k < 6 \sum_{k=1}^{n} m^*(\Omega_k^*) \leqq 6m^*(\Omega^*)$, so that δ_∞ must be zero. Consequently we have $m^*(\Omega_\infty^*) = 0$, and this completes the proof of Theorem 4.

Remark. It is not always true that a proper measurable flow is isomorphic to a flow built under a function. Consider, for example, a generalized flow built under a function $\{T_t\}$ defined as follows: base measure space Ω is the set of all positive integers ($n = 1, 2, \cdots$) on which a measure m is defined for every subset and is equal to the number of points (hence $m(\Omega) = +\infty$); base transformation S is the identity; and the ceiling function is defined by $f(n) = 2^{-n}$. It is clear that $\{T_t\}$ is not isomorphic to a flow built under a function.

6. **Isomorphism to continuous flows.** In this section we prove that every measurable flow on a certain kind of measure space—and this class includes most of the measure spaces usually considered—is isomorphic to a continuous flow on a separable metric space, on which the measure and topology are nicely related. When a measurable flow is given, it is trivial to find a metric with respect to which the flow is continuous, but this metric may not be related with the measure in any way.[17] The point of this isomorphism theorem is to show that it can be done with the measure and the topology properly related.

We begin with some definitions and lemmas.

DEFINITION 10. A measure space $\Omega(\mathfrak{B}, m)$ is *properly separable*,[18] if there exists a countable collection $\mathfrak{A}$ of measurable sets such that the Borel field determined by $\mathfrak{A}$, when completed with respect to the measure m, is exactly the collection $\mathfrak{B}$ of all measurable sets in Ω. $\mathfrak{A}$ is called a *basis* of $\Omega(\mathfrak{B}, m)$.

A countable collection $\mathfrak{A}$ of measurable sets is clearly a basis of a measure space $\Omega(\mathfrak{B}, m)$ if it satisfies the following condition:

(6.1) For any measurable set $M \in \mathfrak{B}$ and for any $\epsilon > 0$, there exists a sequence of measurable sets $\{M_n\}$ ($M_n \in \mathfrak{A}$, $n = 1, 2, \cdots$) such that $M \subset \sum_{n=1}^{\infty} M_n$ and $m\left(\sum_{n=1}^{\infty} M_n\right) < m(M) + \epsilon$. Moreover, a necessary and sufficient condition that

[17] For example, we might define a distance by $d(\omega, v) = 1$ if ω and v are on different trajectories, and $d(\omega, v) = \min(1, |t|)$ if ω and v are on the same trajectory and $\omega_t = v$.

[18] The conditions of proper separability and the existence of a separating sequence are logically independent.

a measure space $\Omega(\mathfrak{B}, m)$ be properly separable is that there exist a countable basis which satisfies the following stronger condition:

(6.2) For any measurable set $M \in \mathfrak{B}$ and for any $\epsilon > 0$, there exists a sequence of measurable sets $\{M_n\}$ $(M_n \in \mathfrak{A},\ n = 1, 2, \cdots)$ such that $M \subset \sum_{n=1}^{\infty} M_n$ and $\sum_{n=1}^{\infty} m(M_n) < m(M) + \epsilon$.

If $\mathfrak{A}$ is a basis of a measure space $\Omega(\mathfrak{B}, m)$, then the field determined by $\mathfrak{A}$ clearly satisfies the condition (6.2).

LEMMA 5. *Let $\Omega^*(\mathfrak{B}^*, m^*)$ be the product measure space of two measure spaces $\Omega(\mathfrak{B}, m)$ and $\Omega'(\mathfrak{B}', m')$. Then $\Omega^*(\mathfrak{B}^*, m^*)$ is properly separable if and only if both $\Omega(\mathfrak{B}, m)$ and $\Omega'(\mathfrak{B}', m')$ are properly separable.*

Proof. We shall first prove that if $\Omega(\mathfrak{B}, m)$ and $\Omega'(\mathfrak{B}', m')$ are both properly separable, then so is $\Omega^*(\mathfrak{B}^*, m^*)$. Let $\mathfrak{A}$ and $\mathfrak{A}'$ be the countable bases of $\Omega(\mathfrak{B}, m)$ and $\Omega'(\mathfrak{B}', m')$ respectively which have the property (6.2). Let us consider the collection $\mathfrak{A}^*$ of all sets M^* of the form: $M^* = M \times M'$, $M \in \mathfrak{A}$, $M' \in \mathfrak{A}'$. Then $\mathfrak{A}^*$ is also countable, and all we have to prove is that $\mathfrak{A}^*$ has the property (6.2). In order to prove this, let $M^* \in \mathfrak{B}^*$ and let $\epsilon > 0$ be an arbitrary positive number. By the definition of the product measure space, there exist two sequences of measurable sets $\{M_n\}$ $(M_n \in \mathfrak{B},\ n = 1, 2, \cdots)$ and $\{M'_n\}$ $(M'_n \in \mathfrak{B}',\ n = 1, 2, \cdots)$ such that $M^* \subset \sum_{n=1}^{\infty} M_n \times M'_n$ and $\sum_{n=1}^{\infty} m(M_n)m'(M'_n) < m^*(M^*) + \frac{\epsilon}{2}$. Since $\mathfrak{A}$ and $\mathfrak{A}'$ both have the property (6.2), there exist, for each n, two sequences of measurable sets $\{M_{n,k}\}$ $(M_{n,k} \in \mathfrak{A},\ k = 1, 2, \cdots)$ and $\{M'_{n,k}\}$ $(M'_{n,k} \in \mathfrak{A}',\ k = 1, 2, \cdots)$ such that

$$M_n \subset \sum_{k=1}^{\infty} M_{n,k}, \qquad \sum_{k=1}^{\infty} m(M_{n,k}) < m(M_n) + \epsilon_n,$$

$$M'_n \subset \sum_{k=1}^{\infty} M'_{n,k}, \qquad \sum_{k=1}^{\infty} m'(M'_{n,k}) < m'(M'_n) + \epsilon_n,$$

where ϵ_n is taken so small that we have $\epsilon_n m(M_n) + \epsilon_n m'(M_n) + \epsilon_n^2 < \frac{\epsilon}{2^{n+1}}$. Consequently we have

$$M^* \subset \sum_{n=1}^{\infty} M_n \times M'_n \subset \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} M_{n,k} \times M'_{n,l},$$

$$\sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} m^*(M_{n,k} \times M'_{n,l}) = \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} m(M_{n,k})m'(M'_{n,l})$$

$$< \sum_{n=1}^{\infty} (m(M_n) + \epsilon_n)(m'(M'_n) + \epsilon_n) < \sum_{n=1}^{\infty} \left(m(M_n)m'(M'_n) + \frac{\epsilon}{2^{n+1}} \right)$$

$$< m^*(M^*) + \epsilon,$$

which proves that the property (6.2) is true for the collection $\mathfrak{A}^*$.

We shall next prove that, conversely, the proper separability of $\Omega^*(\mathfrak{B}^*, m^*)$ implies that of $\Omega(\mathfrak{B}, m)$ and $\Omega'(\mathfrak{B}', m')$. Let $\mathfrak{A}^* = \{M_n^*\}$ $(M_n^* \in \mathfrak{B}^*, n = 1, 2, \cdots)$ be a basis of $\Omega^*(\mathfrak{B}^*, m^*)$ which has the property (6.2). By the definition of the product measure space, there exist, for each M_n^* and for each positive integer p, two sequences of measurable sets $\{M_{n,p,k}\}$ $(M_{n,p,k} \in \mathfrak{B}, k = 1, 2, \cdots)$ and $\{M'_{n,p,k}\}$ $(M'_{n,p,k} \in \mathfrak{B}', k = 1, 2, \cdots)$ such that

$$M_n^* \subset \sum_{k=1}^{\infty} M_{n,p,k} \times M'_{n,p,k}, \tag{6.3}$$

$$\sum_{k=1}^{\infty} m(M_{n,p,k}) m'(M'_{n,p,k}) < m^*(M_n^*) + \frac{1}{p}. \tag{6.4}$$

We shall show that the countable collection $\mathfrak{A} = \{M_{n,p,k}\}$ $(n, p, k = 1, 2, \cdots)$ has the property (6.1) in the measure space $\Omega(\mathfrak{B}, m)$.

In order to prove this, let $M \in \mathfrak{B}$, let ϵ be an arbitrary positive number, and let $\epsilon' = \epsilon \cdot m'(\Omega')$. Let us consider the set $M^* = M \times \Omega' \subset \Omega^*$. M^* clearly belongs to $\mathfrak{B}^*$. Since the collection $\mathfrak{A}^* = \{M_n^*\}$ $(n = 1, 2, \cdots)$ has the property (6.2), there exists a subsequence $\{M_{n_i}^*\}$ $(i = 1, 2, \cdots)$ such that $M^* \subset \sum_{i=1}^{\infty} M_{n_i}^*$ and $\sum_{i=1}^{\infty} m^*(M_{n_i}^*) < m^*(M^*) + \frac{\epsilon'}{2}$. By putting $n = n_i$ and taking p_i larger than $\frac{2^{i+1}}{\epsilon'}$ in (6.3) and (6.4), we have

$$M^* \subset \sum_{i=1}^{\infty} M_{n_i}^* \subset \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} M_{n_i,p_i,k} \times M'_{n_i,p_i,k},$$

$$m^*(M^*) \leqq \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} m(M_{n_i,p_i,k}) m'(M'_{n_i,p_i,k})$$

$$< \sum_{i=1}^{\infty} \left(m^*(M_{n_i}^*) + \frac{\epsilon'}{2^{i+1}} \right) < m^*(M^*) + \epsilon',$$

or $m^*(\Lambda^*) < \epsilon'$ if we put $\Lambda^* = \sum_{i=1}^{\infty} \sum_{k=1}^{\infty} M_{n_i,p_i,k} \times M'_{n_i,p_i,k} - M^*$. Consequently, by Fubini's theorem, there exists a point $\omega' \in \Omega'$ such that $m(\Lambda^*(\omega')) < \epsilon$, where $\Lambda^*(\omega')$ is the set of all points $\omega \in \Omega$ such that $\omega^* = (\omega, \omega') \in \Lambda^*$. Hence, if we denote by $\sum_{i,k}^*$ the sum over all pairs of i and k for which $\omega' \in M'_{n_i,p_i,k}$, then we have $M \subset \sum_{i,k}^* M_{n_i,p_i,k}$ and $m(\sum_{i,k}^* M_{n_i,p_i,k}) < m(M) + \epsilon$, which proves that the property (6.1) holds for the collection $\mathfrak{A}$. Since the same thing can be proved for the collection $\mathfrak{A}'$, the proof of Lemma 5 is completed.

LEMMA 6. *Let $\Omega^*(\mathfrak{B}^*, m^*)$ be the product measure space of two measure spaces $\Omega(\mathfrak{B}, m)$ and $\Omega'(\mathfrak{B}', m')$. If $\Omega^*(\mathfrak{B}^*, m^*)$ has a separating sequence of measurable sets, then there exist two sets of measure zero $N \subset \Omega$ and $N' \subset \Omega'$ such that each of the remaining parts $\Omega - N$ and $\Omega' - N'$ has a separating sequence of measurable sets.*

Proof. Let $\{M_n^*\}$ $(M_n^* \in \mathfrak{B}^*,\ n = 1, 2, \cdots)$ be the separating sequence in Ω^*. For each M_n^* and for each positive integer p, there exist two sequences of measurable sets $\{M_{n,p,k}\}$ $(M_{n,p,k} \in \mathfrak{B},\ k = 1, 2, \cdots)$ and $\{M'_{n,p,k}\}$ $(M'_{n,p,k} \in \mathfrak{B}',\ k = 1, 2, \cdots)$ such that

$$M_n^* \subset \sum_{k=1}^{\infty} M_{n,p,k} \times M'_{n,p,k} \equiv M_{n,p}^*,$$

$$\sum_{k=1}^{\infty} m(M_{n,p,k}) m'(M'_{n,p,k}) < m^*(M_n^*) + \frac{1}{p}.$$

Let us put $N_n^* = \prod_{p=1}^{\infty} M_{n,p}^* - M_n^*$, $n = 1, 2, \cdots$. Then it is clear that $m^*(N_n^*) = 0$, $n = 1, 2, \cdots$. Hence, by Fubini's theorem, there exists a point $\omega' \in \Omega'$ such that $m(N_n^*(\omega')) = 0$, $n = 1, 2, \cdots$, where we denote by $N_n^*(\omega')$ the set of all points $\omega \in \Omega$ for which $\omega^* = (\omega, \omega') \in N_n^*$. We shall prove that if $N = \sum_{n=1}^{\infty} N_n^*(\omega')$, then the sets $\{M_{n,p,k} \cdot (\Omega - N)\}$ $(n, p, k = 1, 2, \cdots)$ form a separating sequence in $\Omega - N$.

Indeed, if there exist two points $\omega, v \in \Omega - N$ which belong to the same members of $\{M_{n,p,k} \cdot (\Omega - N)\}$ $(n, p, k = 1, 2, \cdots)$, then the two points $\omega^* = (\omega, \omega')$ and $v^* = (v, \omega')$ must also belong to the same members of $\{M_{n,p,k} \times M'_{n,p,k}\}$ $(n, p, k = 1, 2, \cdots)$, and consequently to the same members of $\{M_{n,p}^*\}$ $(n, p = 1, 2, \cdots)$. Since ω^* and v^* do not belong to N_n^* for $n = 1, 2, \cdots$, the same thing must be true for the sequence $\{M_n^*\}$ $(n = 1, 2, \cdots)$, which is a contradiction to the separation property of $\{M_n^*\}$ $(n = 1, 2, \cdots)$. Hence $\{M_{n,p,k} \cdot (\Omega - N)\}$ $(n, p, k = 1, 2, \cdots)$ is a separating sequence in $\Omega - N$. Since the same thing can be proved for the measure space $\Omega'(\mathfrak{B}', m')$, the proof of Lemma 6 is completed.

LEMMA 7. *Let $\bar{\Omega}(\overline{\mathfrak{B}}, \bar{m})$ be a properly separable measure space which has a separating sequence of measurable sets, and let $\{T_t\}$ be a measurable flow defined on it. Then there exists an invariant null set $\bar{N} \subset \bar{\Omega}$ and a sequence of measurable sets $\{\bar{M}_n\}$ $(\bar{M}_n \in \overline{\mathfrak{B}},\ n = 1, 2, \cdots)$ contained in $\bar{\Omega}_0 = \bar{\Omega} - \bar{N}$ such that*

(6.5) *$\{\bar{M}_n\}$ $(n = 1, 2, \cdots)$ is at the same time a separating sequence and a basis for the measure space $\bar{\Omega}_0(\overline{\mathfrak{B}}, \bar{m})$,*

(6.6) *if we denote by $\varphi_n(\bar{\omega})$ the characteristic function of the set $\bar{M}_n$, then $\varphi_n(\bar{\omega}_t)$ is a Lebesgue measurable t-function for each $\bar{\omega} \in \bar{\Omega}_0$, and*

$$\frac{1}{\epsilon} \int_0^{\epsilon} \varphi_n(\bar{\omega}_t)\, dt \to \varphi_n(\bar{\omega})$$

as $\epsilon \to 0$ for all $\bar{\omega} \in \bar{\Omega}_0$ (without any exception).

Proof. By Theorem 1, $\bar{\Omega}$ is divided into two invariant measurable sets $\bar{\Omega}_1$ and $\bar{\Omega}_2$ in such a way that $\{T_t\}$ is completely improper on $\bar{\Omega}_1$ and proper on $\bar{\Omega}_2$. It is clear that the measure spaces $\bar{\Omega}_1(\overline{\mathfrak{B}}, \bar{m})$ and $\bar{\Omega}_2(\overline{\mathfrak{B}}, \bar{m})$ are both properly

separable and that each of them has a separating sequence of measurable sets. Hence it is sufficient to prove this lemma only in the two special cases, namely, when $\{T_t\}$ is completely improper and when it is proper.

In the first case we know, by Theorem 2, that $\{T_t\}$ is the identity. Hence we can take as $\{\bar{M}_n\}$ $(n = 1, 2, \cdots)$ the union of the basis and the separating sequence of measurable sets of $\bar{\Omega}_1(\bar{\mathfrak{B}}, \bar{m})$.

In order to prove our lemma in the second case, we may assume, by Theorem 4, that $\{T_t\}$ is a generalized flow built under a function, and since (in Theorem 4) such a flow was obtained by putting together a countable number of flows built under a function, we have only to consider the case when $\{T_t\}$ itself is a flow built under a function. Hence we assume that $\{T_t\}$ is a flow built under a function, and throughout the remainder of this proof, we adopt the notation of Definition 6 $(\bar{\Omega} = \bar{\Omega}_2)$. Moreover, by the proof of Theorem 4, we may assume that there exists a positive constant $c > 0$ such that $f(\omega) > c$ for all $\omega \in \Omega$.

We shall first show that there exists in Ω a null set $N \in \mathfrak{B}$ which is invariant under S and also a sequence $\{M_n\}$ $(M_n \in \mathfrak{B}, n = 1, 2, \cdots)$ of measurable sets contained in $\Omega_0 = \Omega - N$ which is at the same time a separating sequence and a basis for a properly separable measure space $\Omega_0(\mathfrak{B}, m)$.

For this purpose, consider a subset $\bar{\Omega}_c$ of $\bar{\Omega}$ defined by

$$\bar{\Omega}_c = [\tilde{\omega}: 0 \leqq G(\tilde{\omega}) < c].$$

Then we have $\bar{\Omega}_c = \Omega \times [0, c)$, where $[0, c)$ is a semi-open interval $0 \leqq u < c$ of real numbers, and it is easy to see that the measure space $\bar{\Omega}_c(\bar{\mathfrak{B}}, \bar{m})$ is the product measure space of $\Omega(\mathfrak{B}, m)$ and the interval $[0, c)$ taken with Lebesgue measure. Hence, by Lemmas 5 and 6, the measure space $\Omega(\mathfrak{B}, m)$ has a countable basis, and after throwing out a null set M of measure zero, the remainder $\Omega - M$ contains a separating sequence of measurable sets. Let us put $N = \sum_{k=-\infty}^{+\infty} S^k(M)$. Then N is an invariant null set, and it is clear that $\Omega_0 = \Omega - N$ has both a basis and a separating sequence of measurable sets.

Now we are ready to obtain the required invariant null set N and the sequence of measurable sets $\{\bar{M}_n\}$ $(n = 1, 2, \cdots)$ with the properties (6.5) and (6.6). We first define $\bar{N}$ as the set of all points $\tilde{\omega}$ of the form $\tilde{\omega} = (\omega, u)$, $\omega \in N$, $0 \leqq u < f(\omega)$. $\bar{N}$ is clearly an invariant null set in $\bar{\Omega}$. Let us further define the sets $\bar{M}_{n,a,b}$ by

$$\bar{M}_{n,a,b} = (M_n \times [a, b))(\bar{\Omega} - \bar{N}),$$

where a and b are two rational numbers such that $0 < a < b < +\infty$, and $[a, b)$ is a semi-open interval $a \leqq u < b$. It is clear that these $\bar{M}_{n,a,b}$ are the required sets. We have only to rearrange them into a simple sequence.

DEFINITION 11. Let $\Omega(\mathfrak{B}, m)$ be a measure space, and let Ω be a topological space at the same time. A measurable set $M \in \mathfrak{B}$ is called *regular* if for any $\epsilon > 0$ there exists an open measurable set $O \in \mathfrak{B}$ such that $M \subset O$ and $m(O) <$

$m(M) + \epsilon$. A measure space is called *regular* if every measurable set $M \in \mathfrak{B}$ is regular.

DEFINITION 12. A measure space $\Omega(\mathfrak{B}, m)$ is an *M-space* if it satisfies the following conditions:

(6.7) Ω is a separable metric space,

(6.8) every open set is measurable and has a positive measure,

(6.9) $\Omega(\mathfrak{B}, m)$ is a regular measure space.

DEFINITION 13. A flow $\{T_t\}$ defined on a topological measure space is called *continuous* if $T_t(\omega)$ is a continuous function of two variables ω and t.

THEOREM 5. *Let $\{T_t\}$ be a measurable flow defined on a measure space which is properly separable and has a separating sequence of measurable sets. Then $\{T_t\}$ is isomorphic to a continuous flow on an M-space.*

Proof. We first apply Lemma 7 and find an invariant null set N and a sequence of measurable sets $\{M_n\}$ $(n = 1, 2, \cdots)$ with the properties (6.5) and (6.6) of the sets $\bar{N}$ and $\{\bar{M}_n\}$ $(n = 1, 2, \cdots)$ of that lemma. (We change to the simpler notation because it will not be necessary here to represent our flow as a flow built under a function.)

Let us denote by $\varphi_n(\omega)$ the characteristic function of the set M_n. For each $\omega \in \Omega_0 = \Omega - N$, define a double sequence of functions $\{x^{\omega}_{n,m}(t)\}$ $(n, m = 1, 2, \cdots)$ by

$$x^{\omega}_{n,m}(t) = m \int_0^{\frac{1}{m}} \varphi_n(\omega_{t+s})\, ds, \qquad -\infty < t < +\infty.$$

It is clear that all these functions are continuous and uniformly bounded: $0 \leqq x^{\omega}_{n,m}(t) \leqq 1$. Let us define the distance $d(\omega, v)$ of two points ω and v of Ω_0 by

$$(6.10) \qquad d(\omega, v) = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \sum_{k=1}^{\infty} 2^{-(n+m+k)} \sup_{|t| \leqq k} |\, x^{\omega}_{n,m}(t) - x^{v}_{n,m}(t)\,|.$$

We shall prove that the following conditions are satisfied:

(6.11) $\Omega_0 = \Omega - N$ is a separable metric space with $d(\omega, v)$ as a distance,

(6.12) all open sets of Ω_0 are measurable,

(6.13) $m(M)$ is a regular measure on Ω_0,

(6.14) $\{T_t\}$ is a continuous flow on Ω_0.

In order to prove (6.11), let Ξ be the space of all sequences of functions $\xi = \{x_{n,m}(t)\}$ $(n, m = 1, 2, \cdots)$, where each $x_{n,m}(t)$ is a real valued con-

tinuous function defined for $-\infty < t < +\infty$ such that $0 \leqq x_{n,m}(t) \leqq 1$. As is well known, Ξ is a separable metric space with respect to the distance

$$d(\xi, \eta) = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} \sum_{k=1}^{\infty} 2^{-(m+n+k)} \sup_{|t| \leqq k} |x_{n,m}(t) - y_{n,m}(t)|, \tag{6.15}$$

where $\xi = \{x_{n,m}(t)\}$, $\eta = \{y_{n,m}(t)\}$ $(n, m = 1, 2, \cdots)$. If we now put[19] $\xi(\omega) = \{x^{\omega}_{n,m}(t)\}$ $(n, m = 1, 2, \cdots)$, then $\omega \to \xi(\omega)$ is a mapping of Ω_0 onto a subset Ξ_0 of Ξ. Since $x^{\omega}_{n,m}(0) \to \varphi_n(\omega)$ as $m \to +\infty$ (this follows from (6.6)), and since $\omega \neq v$ $(\omega, v \in \Omega_0)$ implies the existence of an n for which $\varphi_n(\omega) \neq \varphi_n(v)$ (this follows from (6.5)), this mapping gives a one-to-one correspondence of Ω_0 and Ξ_0. Since (6.10) and (6.15) give the same distance for corresponding points, and since, as a subset of a separable metric space Ξ, Ξ_0 itself is also metric and separable, the proof of (6.11) is completed.

Since (6.11) is already proved, in order to prove (6.15), it is sufficient[20] to show that every sphere in Ω_0 (defined by the metric $d(\omega, v)$) is measurable, and this will be done if we can prove that for each fixed v, the distance $d(\omega, v)$ is a measurable function of ω. By the definition (6.10) of $d(\omega, v)$, it is sufficient to show that, for each fixed v, the ω-function $\psi_{n,m}(\omega) = \sup_{|t| \leqq k} |x^{\omega}_{n,m}(t) - x^{v}_{n,m}(t)|$ is measurable. This is, however, a direct consequence of the following two facts: (1) since $x^{\omega}_{n,m}(t)$ and $x^{v}_{n,m}(t)$ are both continuous in t, the sup in the formula above may be replaced by the sup for all rational numbers t which satisfy $|t| \leqq k$; (2) for every fixed t, $x^{\omega}_{n,m}(t)$ is a measurable function of ω.

Now we proceed to the proof of (6.13). We shall prove that for any $M \in \mathfrak{B}$ there exists a sequence of open sets $\{O_k\}$ $(k = 1, 2, \cdots)$ such that $M \subset O_k$ $(k = 1, 2, \cdots)$ and $m(O_k - M) \to 0$. Since the measure space $\Omega(\mathfrak{B}, m)$ is properly separable with $\{M_n\}$ $(n = 1, 2, \cdots)$ as a basis, it is sufficient to show this for each M_n. In order to prove that each M_n is regular, we define the sets $O_{n,k}$ by

$$O_{n,k} = \sum_{m=k}^{\infty} [\omega : x^{\omega}_{n,m}(0) > \tfrac{1}{2}], \qquad k = 1, 2, \cdots.$$

Then it is clear that each $O_{n,k}$ is open and that we have $M_n = \prod_{k=1}^{\infty} O_{n,k}$. (This means that M_n itself is a G_δ-set.) The proof of (6.13) is completed.

Lastly, we have to prove (6.14). This can be proved easily by a standard method, using the fact that

$$|x^{\omega}_{n,m}(t) - x^{\omega}_{n,m}(s)| \leqq 2m\,|t - s|,$$

which is clear from (5.4) and the definition of $x^{\omega}_{n,m}(t)$. We omit the proof.

[19] The idea of mapping a flow into the space of functions of real variables in such a way that the given flow goes into the translation flow on function space has been exploited by J. L. Doob in [2], Theorem 10, p. 769.

[20] This is sufficient since in a separable metric space every open set is a sum of a countable number of spheres.

Thus we have proved (6.11), (6.12), (6.13) and (6.14). Now, let N_0 be the set of all points $\omega \in \Omega_0$, such that there exists a sphere of measure zero containing this point. Then N_0 is an open subset of Ω_0, and since by the separability of Ω_0, N_0 is expressed as a sum of a countable number of such spheres, N_0 itself must be of measure zero. Moreover, since the flow $\{T_t\}$ is continuous on Ω_0, N_0 must be an invariant set. Consequently, if we consider the space $\Omega_{00} = \Omega_0 - N_0 = \Omega - N - N_0$, then Ω_{00} is an M-space and $\{T_t\}$ is a continuous flow on Ω_{00}. Since the excluded set $N + N_0$ is an invariant null set, the proof of Theorem 5 is completed.

Bibliography

0. W. Ambrose, *Change of velocities in a continuous ergodic flow*, Duke Mathematical Journal, vol. 8(1941), pp. 425–440.
1. W. Ambrose, *Representation of ergodic flows*, Annals of Mathematics, vol. 42(1941), pp. 723–739.
2. J. L. Doob, *One-parameter families of transformations*, Duke Mathematical Journal, vol. 4(1938), pp. 752–774.
3. E. Hopf, *Ergodentheorie*, Berlin, 1937.
4. J. von Neumann, *Zur Operatorenmethode in der klassischen Mechanik*, Annals of Mathematics, vol. 33(1932), pp. 587–642.
5. S. Saks, *Theory of the Integral*, Warsaw, 1937.
6. N. Wiener, *The ergodic theorem*, Duke Mathematical Journal, vol. 5(1939), pp. 1–18.

Institute for Advanced Study.

THE DECOMPOSITION OF MEASURES, II

By Warren Ambrose, Paul R. Halmos, and Shizuo Kakutani

The main purpose of this paper is to prove that a flow on a measure space may be split into ergodic parts. The first result of this type is due to von Neumann,[1] who worked with metric, complete, separable spaces. For some important applications, particularly in probability theory, it is necesary to dispense with these topological assumptions. We shall make extensive use of the terminology, notation, and results of (D)[2] and (S).[3] Incidentally, we find it necessary to relax somewhat the strict separability conditions of (D), redefine "direct sum", and prove a generalization of the general decomposition theorem of (D). As for the decomposition of a flow, we have chosen to reduce this to the case of a single measure preserving transformation by means of Th rem 4 of (S).

A measure space $\Omega(\mathfrak{M}, m)$ is *properly separable* if there exists a strictly separable Borel field $\mathfrak{B} \subseteq \mathfrak{M}$, such that for every $M \in \mathfrak{M}$ there is a $B \in \mathfrak{B}$ with $M \subseteq B$ and $m(B - M) = 0$. Throughout this paper we assume that all measure spaces considered are properly separable and complete, in the sense that any subset of a measurable set of measure zero is itself measurable. Ω is said to be a *direct sum* of the measure spaces $Y_x(\mathfrak{Y}_x, \nu_x)$ formed with respect to the measure space $X(\mathfrak{X}, \mu)$, in symbols

$$\Omega(\mathfrak{M}, m) = \int_{X(\mathfrak{X}, \mu)} Y_x(\mathfrak{Y}_x, \nu_x)\, d\mu(x),$$

if the conditions of §3 in (D) are satisfied, with the exception that we require only that for every $M \in \mathfrak{M}$, MY_x be a measurable subset of Y_x for *almost every* x.

Theorem 1. *If $\Omega(\mathfrak{M}, m)$ is a measure space and $\mathfrak{A}$ a Borel field, $\mathfrak{A} \subseteq \mathfrak{M}$, then there exists a set $A \in \mathfrak{A}$ of measure zero such that $\Omega - A$ is a direct sum,*

$$\Omega - A = \int_X Y_x\, d\mu(x),$$

in such a way that the Borel field $\mathfrak{X}$ of all measurable x-sets is contained in and is equivalent to the given Borel field $\mathfrak{A}$.

Proof. Let $\mathfrak{B}$ be a strictly separable Borel field related to $\mathfrak{M}$ as in the definition of proper separability, and let $\mathfrak{A}'$ be any strictly separable Borel field con-

Received June 13, 1941.

[1] J. v. Neumann, *Zur Operatorenmethode in der klassischen Mechanik*, Annals of Mathematics, vol. 33(1932), pp. 587–642. See p. 617.

[2] (D): Paul R. Halmos, *The decomposition of measures*, Duke Mathematical Journal, vol. 8(1941), pp. 386–392.

[3] (S): W. Ambrose and S. Kakutani, *Structure and continuity of measurable flows*, Duke Mathematical Journal, vol. 9(1942), pp. 25–42.

tained in $\mathfrak{A}$ and equivalent to it.[4] We may take $\mathfrak{A}' \subseteq \mathfrak{B}$. Now for the moment we consider only the Borel fields $\mathfrak{B}$ and $\mathfrak{A}'$ in Ω, and apply Theorem 1 of (D); this tells us that there exists a set $A \in \mathfrak{A}'$, and therefore $A \in \mathfrak{A}$, of measure zero such that

$$\Omega(\mathfrak{B}, m) - A = \int_{X(\mathfrak{X}', \mu)} Y_x(\mathfrak{Y}_x', \nu_x)\, d\mu(x),$$

and such that the Borel field $\mathfrak{X}'$ of all measurable x-sets coincides with $\mathfrak{A}'$ in $\Omega - A$. Let $\mathfrak{X}$ and $\mathfrak{Y}_x$ be the Borel fields obtained from $\mathfrak{X}'$ and $\mathfrak{Y}_x'$ respectively, by adjoining all subsets of sets of measure zero. We shall prove that

$$\Omega(\mathfrak{M}, m) - A = \int_{X(\mathfrak{X}, \mu)} Y_x(\mathfrak{Y}_x, \nu_x)\, d\mu(x).$$

In other words we shall prove that for each $M \in \mathfrak{M}$, $MY_x \in \mathfrak{Y}_x$ for almost all x, and that the integral formula

$$m(M) = \int_X \nu_x(MY_x)\, d\mu(x) \tag{1}$$

is valid. From the definition of proper separability we know that there exist two sets B_1 and B_2 in $\mathfrak{B}$ with $B_1 \subseteq M \subseteq B_2$ and $m(B_2 - B_1) = 0$. Moreover, B_1Y_x and B_2Y_x are measurable (in Y_x) for all x, $\nu_x(B_1Y_x)$ and $\nu_x(B_2Y_x)$ are measurable functions of x, and the integral formula (1) is valid with B_1 or B_2 in place of M. This implies that $\nu_x(B_1Y_x) = \nu_x(B_2Y_x)$ for almost all x, so that for almost all x, MY_x is measurable and

$$\nu_x(B_1Y_x) = \nu_x(MY_x) = \nu_x(B_2Y_x).$$

Since $m(B_1) = m(M) = m(B_2)$, this implies that (1) is valid for all $M \in \mathfrak{M}$.

We can apply Theorem 1 to prove a decomposition theorem for measure preserving transformations which differs in two ways from Theorem 2 of (D). It applies to the case where the space is properly separable rather than strictly separable, and it does not require any separability assumptions on the collection of invariant sets. The first of these differences is unimportant and merely makes the theorem a little more general, but the second is essential for any effective applications since in general the invariant sets will not form a strictly separable (nor even a properly separable) Borel field. The proof is essentially the same as the proof of Theorem 2 of (D), except that a more ingenious argument (due to von Neumann) is necessary in order to prove ergodicity on the spaces Y_x which we obtain. For this purpose we shall make use of the following simple consequence of the Birkhoff ergodic theorem.[5] A necessary and sufficient condition that a measure preserving transformation T on a measure space $\Omega(\mathfrak{M}, m)$ be ergodic is that for some sequence $B_1, B_2, \cdots$ of measurable sets, which span a Borel field be equivalent to $\mathfrak{M}$,

$$\frac{1}{N} \sum_{k=0}^{N-1} \varphi(B_n, T^k \omega)$$

[4] See Lemma 1, (D). To obtain $\mathfrak{A}' \subseteq \mathfrak{B}$ we may replace $\mathfrak{B}$ by the Borel field spanned by $\mathfrak{A}'$ and $\mathfrak{B}$.

[5] E. Hopf, *Ergodentheorie*, Berlin, 1937. See pp. 49–54.

converge almost everywhere, as $N \to \infty$, to a constant independent of ω, (where $\varphi(B, \omega)$ is the characteristic function of B).

THEOREM 2. *If T is a measure preserving transformation, then there exists an invariant set A of measure zero such that $\Omega - A$ is a direct sum,*

$$\Omega - A = \int_X Y_x \, d\mu(x),$$

in such a way that each Y_x is invariant under T and T is an ergodic measure preserving transformation on Y_x.

Proof. Let $\mathfrak{C}$ be the Borel field of invariant sets and apply Theorem 1 to $\Omega(\mathfrak{M}, m)$ and $\mathfrak{C}$. By a slight modification of the corresponding part of the proof of Theorem 2 in (D) we prove that T on Y_x is measure preserving for almost all x. We shall explicitly prove only the ergodicity.

Let $B_1, B_2, \cdots$ be a sequence of measurable sets which span a Borel field $\mathfrak{B}$ equivalent to $\mathfrak{M}$. We shall prove that for almost every fixed x,

$$\frac{1}{N} \sum_{k=0}^{N-1} \varphi(B_n Y_x, T^k \omega)$$

converges to a constant independent of ω. Since we know, from the ergodic theorem, that

$$\frac{1}{N} \sum_{k=0}^{N-1} \varphi(B_n, T^k \omega)$$

does converge almost everywhere (with respect to the measure m) to a measurable function $f_n(\omega)$, this amounts to proving that $f_n(\omega)$ depends on x alone. Since $f_n(\omega)$ is invariant under T, $f_n(T\omega) = f_n(\omega)$, we know that $f_n(\omega)$ is measurable $(\mathfrak{C})$. Since $\mathfrak{X}$ is equivalent to $\mathfrak{C}$, we may find functions $f_n'(\omega)$, which are measurable $(\mathfrak{X})$, so that $f_n(\omega) = f_n'(\omega)$ almost everywhere. Increasing the set A of measure zero, described in Theorem 1, so as to include all points in every Y_x on which T is not measure preserving, and all points ω for which $f_n(\omega)$ is not defined, or for which $f_n(\omega) \neq f_n'(\omega)$, $n = 1, 2, \cdots$, we obtain the desired result.

The ergodic theorem can also be used to prove not only the ergodicity but also the existence of the decomposition. For every measurable set B the averages

$$\frac{1}{N} \sum_{k=0}^{N-1} \varphi(B, T^k \omega)$$

converge almost everywhere to a function $\beta(B, \omega)$ (depending, of course, on B), which is invariant under T. Moreover, for any invariant set A, we have

$$\int_A \beta(B, \omega) \, dm(\omega) = \lim_{N \to \infty} \frac{1}{N} \sum_{k=0}^{N-1} \int_A \varphi(B, T^k \omega) \, dm(\omega) = m(AB).$$

(The integrability of β and the legitimacy of term by term integration are easily established.) A comparison with §4 of (D) shows that $\beta(B, \omega)$ has the defining

properties of the $\delta(B, \omega)$ described there. Consequently, $\beta(B, \omega) = \delta(B, \omega)$ almost everywhere. The chief difficulty in the proof of Theorem 1 in (D) was the application of the Radon-Nikodym differentiation theorem to prove the existence of $\delta(B, \omega)$: the equality just proved suggests another approach, by invoking the ergodic theorem and defining $\delta(B, \omega)$ to be the limit of the averages discussed above. This approach leads to the same technical difficulties as before in connection with the proof that for fixed ω, $\delta(B, \omega)$ is a measure on $\mathfrak{M}$, but is otherwise a little simpler than the proof using the Nikodym theorem. It has the disadvantage of applying to Theorem 2 only; the differentiation seems to be indispensable in the proof of Theorem 1.

THEOREM 3. *If T_t is a measurable flow on a measure space $\bar{\Omega}(\overline{\mathfrak{M}}, \bar{m})$ which contains a separating sequence of measurable sets,*[6] *then there exists an invariant set $\bar{A}$ of measure zero such that $\bar{\Omega} - \bar{A}$ is a direct sum,*

$$\bar{\Omega} - \bar{A} = \int_X \bar{Y}_x \, d\mu(x),$$

in such a way that each $\bar{Y}_x$ is invariant under T_t and T_t is an ergodic measurable flow on $\bar{Y}_x$.

Proof. By Theorems 1, 2, and 4 of (S) it is possible to split $\bar{\Omega}$ into two subspaces. For one of these the flow is the identity (for which this theorem is trivial). The other can be split into a countable sum of invariant measure spaces on each of which the flow is (isomorphic to) a flow built under a function. Hence, in the remainder of this proof, we shall assume that T_t is built under a function and we shall use without explanation all the notation of Definition 6 in (S).[7] We shall prove this theorem for T_t by applying Theorem 2 to the transformation S on Ω, but for this purpose we must know that Ω is properly separable; this follows from Lemma 4 of (S) and the proper separability of $\bar{\Omega}$.

Applying Theorem 2 to S we find a set A of measure zero invariant under S such that

$$\Omega - A = \int_{X(\mathfrak{X}, \mu)} Y_x(\mathfrak{Y}_x, \nu_x) \, d\mu(x), \tag{2}$$

and such that Y_x is invariant under S and S is ergodic on Y_x. We define the function $f_x(\omega)$ on Y_x by $f_x(\omega) = f(\omega)$ for $\omega \in Y_x$, and we define the space $\bar{Y}_x$ by $\bar{Y}_x = \bar{\Omega}(Y_x \times R)$, (where R denotes the real line); we make $\bar{Y}_x$ into a measure space, $\bar{Y}_x = \bar{Y}_x(\overline{\mathfrak{Y}_x}, \bar{\nu}_x)$, by defining measure multiplicatively in terms of ν_x on Y_x and Lebesgue measure on R. It follows from the definition of a direct sum that $f_x(\omega)$ is measurable $(\mathfrak{Y}_x)$, except possibly for an x-set, X_0, of measure zero. We define $\bar{A}$ by

[6] See Definition 8 in (S). This theorem could be stated more generally, without assuming the existence of a separating sequence, but the proof would then involve a consideration of improper flows, (see Definition 7 of (S)). The present case is the only one of interest.

[7] We change one detail in the notation of (S) by writing $\bar{\Omega}(\overline{\mathfrak{M}}, \bar{m})$ and $\Omega(\mathfrak{M}, m)$ in place of $\bar{\Omega}(\overline{\mathfrak{B}}, \bar{m})$ and $\Omega(\mathfrak{B}, m)$.

$$\bar{A} = \bar{\Omega}(A \times R) + \sum_{x \in X_0} \bar{Y}_x ;$$

we shall show that

$$\bar{\Omega} - \bar{A} = \int_{X-X_0} \bar{Y}_x \, d\mu(x), \tag{3}$$

and that this decomposition has all the properties asserted in the theorem.

It is clear, since S leaves A and each Y_x invariant, that T_t leaves $\bar{A}$ and each $\bar{Y}_x$ invariant; also, from the proper separability of each $Y_x(\mathfrak{Y}_x , \nu_x)$ we conclude that each $\bar{Y}_x(\overline{\mathfrak{Y}}_x , \bar{\nu}_x)$ is properly separable. We know that, on each Y_x , S is an ergodic measure preserving transformation, and that (for $x \in X - X_0$) $f_x(\omega)$ is measurable $(\mathfrak{Y}_x)$; since T_t , when considered only on $\bar{Y}_x$, is the flow built under $f_x(\omega)$ on the transformation S, it follows that T_t is an ergodic measurable flow[8] on each $\bar{Y}_x$, $(x \in X - X_0)$.

To complete the proof we need only establish the direct sum relation (3). For this purpose let $\overline{\mathfrak{M}}_0$ be the collection of all sets $\bar{M} \subseteq \bar{\Omega} - \bar{A}$ satisfying the following conditions:

(i) $\bar{M} \in \overline{\mathfrak{M}}$;

(ii) $\bar{M}\bar{Y}_x \in \overline{\mathfrak{Y}}_x$ for almost every $x \in X - X_0$;

(iii) $\bar{\nu}_x(\bar{M}\bar{Y}_x)$ is measurable $(\mathfrak{X})$;

(iv) $\bar{m}\bar{M} = \int_{X-X_0} \bar{\nu}_x(\bar{M}\bar{Y}_x) \, d\mu(x)$.

We shall show that $\overline{\mathfrak{M}}_0 = \overline{\mathfrak{M}}$.

If $\bar{M} \subseteq \bar{\Omega} - \bar{A}$ is of the form

$$\bar{M} = \{\bar{\omega}: F(\bar{\omega}) > c\}\{\bar{\omega}: a \leqq G(\bar{\omega}) < b\}(M \times R), \tag{4}$$

where $M \in \mathfrak{M}$, and $0 \leqq a \leqq b \leqq c$, then it is trivial (from the direct sum relation (2) and the multiplicative definition of the measures $\bar{m}$ and $\bar{\nu}_x$) that $\bar{M} \in \overline{\mathfrak{M}}_0$. It is also easy to see that if $\bar{M} \subseteq \bar{\Omega} - \bar{A}$ is of the form

$$\bar{M} = \{\bar{\omega}: G(\bar{\omega}) > c\}(M \times R), \tag{5}$$

where $M \in \mathfrak{M}$, then $\bar{M}$ is a countable sum of disjoint sets of the form (4), and hence that $\bar{M} \in \overline{\mathfrak{M}}_0$. Since every set in the field determined by sets of the form (4) or (5) is a finite sum of pairwise disjoint sets of this form, it follows that every set in this field is in $\overline{\mathfrak{M}}_0$. Since, finally, $\overline{\mathfrak{M}}_0$ is clearly a normal class, and since along with a set of measure zero $\overline{\mathfrak{M}}_0$ contains all of its subsets, it follows that $\overline{\mathfrak{M}}_0 = \overline{\mathfrak{M}}$.[9]

THE INSTITUTE FOR ADVANCED STUDY.

[8] We use here the following easily proved statement: if T_t is a flow built under a function then the measurability of the function implies the measurability of the flow. It is trivial that a flow built under a function is ergodic if and only if the transformation on the base space is ergodic.

[9] We use here the theorem (see S. Saks, *Theory of the Integral*, Warsaw, 1937, p. 83) that the normal class determined by a field coincides with the Borel field determined by it.

THE FUCHSIAN EQUATION OF SECOND ORDER WITH FOUR SINGULARITIES

By A. Erdélyi

1. In this paper the solutions of the Fuchsian equation of second order with four singularities are investigated by means of series of hypergeometric functions.

A linear differential equation of second order with four singularities which are "regular points" (this name is due to Thomé whereas Fuchs himself used the term "points of determinateness"; both names appear to be rather inadequate) can be reduced by a linear transformation of the variables to the equation defined by the Riemannian scheme

$$P\left\{\begin{matrix} 0 & 1 & a & \infty & \\ 0 & 0 & 0 & \alpha & x \\ 1-\gamma & 1-\delta & 1-\epsilon & \beta & \end{matrix}\right\}, \tag{1.1}$$

where the exponents are connected by Riemann's relation

$$\alpha+\beta-\gamma-\delta-\epsilon+1=0. \tag{1.2}$$

Heun's equation defined by the scheme (1.1) is of considerable theoretical interest, for it is the simplest equation of Fuchsian type the coefficients of which are not determined uniquely by the singularities and the exponents attached to the singularities. In fact, in Heun's equation (3.1) there is a constant h which is quite arbitrary from the point of view of the scheme (1.1) and is thus an accessory parameter according to the terminology of F. Klein. From the practical point of view, Heun's equation is of some interest, for many of the differential equations occurring in the applications of analysis are special or limiting cases of Heun's equation. It is sufficient to recall that the hypergeometric and confluent hypergeometric equations, the differential equations of Lamé, Mathieu, Legendre, Bessel and Weber, those of the polynomials of Jacobi, Tchebicheff, Laguerre and Hermite as well as that of Bateman's k-function belong to this class.

2. The simplest way of representing the fundamental branches of the functions defined by the scheme (1.1) is to try power-series. These represent the functions in a circle with one singularity on the boundary and another singularity at the center of the domain of convergence. An alternative plan, proposed in this paper, consists in expanding the solutions of Heun's equation into certain series of hypergeometric functions. In this way, in a certain sense, three singularities may be taken into consideration. The series are convergent in a domain the boundary of which is an elliptic limaçon with two singularities as foci and a third singularity on the circumference. Though these series may offer some points of interest even in the general case (i.e., with arbitrary values

Received June 20, 1941.

of h), yet it is only in a certain important exceptional case that their usefulness is fully revealed.

It is well known that Heun's equation has a solution regular at *two* singularities of the differential equation, if h has certain values satisfying one of a set of transcendental equations. Following the corresponding usage in the case of the equations of Lamé and Mathieu, it seems appropriate to use the term Heun function (of the first kind) for a solution which is regular at two of the four points $0, 1, a, \infty$.

Heun functions have been the subject of several investigations. Lambe and Ward derived integral equations for such functions. Actually, Lambe and Ward suppose $1 - \alpha$ or $1 - \beta$ to be a positive integer. This causes the Heun functions to be regular at *three* singularities (Heun polynomials). This restriction is however by no means necessary; similar theorems hold for *transcendental* Heun functions, with arbitrary values of $1 - \alpha$ and $1 - \beta$. Svartholm investigated the expansion of Heun functions in series of Jacobi polynomials. His series are convergent in an ellipse the foci of which are the two singularities of Heun's equation in which the Heun function is regular. One of the remaining two singularities is on the circumference of the ellipse, except when $1 - \alpha$ or $1 - \beta$ is a positive integer when we have a terminating series representing a Heun polynomial in the whole plane.

Leaving aside Heun *polynomials* for the moment, we see that *transcendental* Heun functions are represented by their power-series in a *circle*, by their Jacobi-series in an *ellipse* of convergence. The series of hypergeometric functions dealt with in this paper is convergent in the exceptional case of Heun functions *in the whole plane* of the complex variable x. The domain of convergence thus including the two singularities of the Heun function, the behaviour of this function at its singularities may be studied in every detail. So far as I am aware, this is the first step towards an explicit knowledge of the monodromic group of Heun's equation.

Another useful feature of our series shows itself when dealing with the general solutions of Heun's equation in the exceptional case. Svartholm's series do not appear to be suitable for representing any other solution than Heun's function of first kind. The power-series and the series of hypergeometric functions, however, work as well as in the general case. Besides, series of hypergeometric functions represent the general solution outside of a certain limaçon. This domain containing two singularities of Heun's equation (namely those at which the transcendental Heun function is singular), the general solution at these singularities can be determined. Hence our knowledge of the monodromic group of Heun's equation may be completed as far as the singularities of Heun's function are concerned. Heun's equation has two more singularities (at which Heun's function is regular); our knowledge of the transformations, related to these singularities, of the monodromic group is less complete.

Let us return now to Heun *polynomials*. These are represented by *terminating* series and hence the question of convergence does not arise at all in this

case with functions of the first kind. Yet here too our series are of considerable advantage. While the power-series representation of the *general* solution is still only convergent in a circle, the representation of the general solution as given by our method consists of a *finite* linear combination of hypergeometric functions and is valid in the whole plane. In this case the knowledge of the monodromic group is complete.

Also there are some interesting relations connecting two different series representing the same Heun function.

3. The results merely outlined in the preceding section are essentially the generalizations of some results on Lamé functions which will be published elsewhere. In the present case too, the type of the series to be used could be inferred from the integral equations satisfied by Heun functions. For the sake of brevity, however, I omit this deduction and start directly assuming the form of the series and verifying that it satisfies Heun's equation if the coefficients satisfy certain recurrence formulas. Also some other results, mentioned only in the preceding section, will not be proved in extenso. The content of the following pages in connection with my more detailed exposition of the corresponding investigations on Lamé functions will certainly enable the reader to clear up to his own satisfaction every detail.

I thought it expedient to restrict myself in the main part of this paper to Heun's equation. I might be allowed some remarks, however, on the general import of the underlying method and on its applications to more involved types of differential equations.

To fix the ideas, let us take a differential equation of second order of Fuchsian type (for the general theory of such equations see, for instance, [3],[1] p. 370) with n singularities. There are $n - 3$ accessory parameters in such an equation. The fundamental solutions may be expressed either as power-series (convergent in a circle) or as a series of hypergeometric functions (convergent inside an elliptic limaçon). The coefficients satisfy $(n - 1)$-term and $(2n - 5)$-term recurrence formulas respectively.

In the general case (i.e., with arbitrary values of the accessory parameters) there is only one singularity inside the domain of convergence. If the accessory parameters satisfy some of a certain set of transcendental equations, then there is at least one solution of the Fuchsian equation regular at two, three or more of the singularities. In this case in which the circle or limaçon of convergence increases correspondingly and alternatively, this regular solution (but *not* the general solution of the differential equation in this exceptional case) is expressible by a series of Jacobi polynomials, see [8], convergent inside a certain ellipse. The most interesting case is that in which the accessory parameters satisfy $n - 3$ independent equations so that they take one of a denumerably infinite set of characteristic values. In this case the differential equation has a solution (of first kind) with only two singularities. While power-series and Svar-

[1] The numbers in brackets refer to the bibliography.

tholm's Jacobi series represent this solution only in a finite part of the complex plane, the series introduced here are convergent in the *whole plane* and give full information about the monodromic group as far as the two singularities of the solution of first kind are concerned. The series into powers or those into Jacobi polynomials are convergent in the whole plane only if the solution of first kind is regular at $n-1$ singularities of the differential equation, and this is only possible if a condition is imposed on the exponents. In the last mentioned exceptional case our series are again more satisfactory, for they represent the *general* solution of the differential equation as a *finite* linear combination of hypergeometric functions. Also they yield the full monodromic group.

The question naturally arises if it is possible to represent by series convergent in the whole plane solutions regular at $n-3, n-4, \cdots, 2$ singularities of a Fuschsian equation with n singularities. The answer is in the affirmative. Such a representation is possible by series of certain functions satisfying a suitably chosen set of Fuchsian equations with $4, 5, \cdots, n-1$ singularities. Though theoretically of some interest, owing to our yet imperfect knowledge of the solutions of Fuchsian equations with $4, 5, \cdots$ singularities, this procedure is not likely to be practicable. The general principle is, however, perfectly clear: representing solutions of Fuchsian equations by series of functions which satisfy Fuchsian equations of the same order with a smaller number of singularities. The usual power-series solution fits into this scheme being its lowest stage, for (cf., for instance, [10], §10.8) powers are the solutions of the Fuchsian equation with two singularities only.

It is hardly necessary to mention that the same principle applies equally to Fuchsian equations of any order.

In the rest of this paper I shall restrict myself, for the sake of brevity in the formulas, to Heun's equation

$$L[y] \equiv x(x-1)(x-a)\frac{d^2y}{dx^2} + \{\gamma(x-1)(x-a) + \delta x(x-a) + \epsilon x(x-1)\}\frac{dy}{dx} + \alpha\beta(x-h)y = 0 \tag{3.1}$$

defined by the scheme (1.1), Riemann's relation (1.2) being supposed to be satisfied.

4. It would seem plausible that some series of the sort

$$\sum c_m P\begin{Bmatrix} 0 & 1 & \infty & \\ m & 0 & \alpha & x \\ \delta-\alpha-\beta-m & 1-\delta & \beta & \end{Bmatrix} \tag{4.1}$$

should be able to render the behavior of the function defined by the scheme (1.1) in a domain including 1 and ∞.

There are six different fundamental branches of the P-function occurring in (4.1) which will be denoted by

$$\varphi_m^1 = \varphi_m^1(\alpha, \beta; \delta; x) = \frac{\Gamma(\alpha - \delta + m + 1)\Gamma(\beta - \delta + m + 1)}{\Gamma(\alpha + \beta - \delta + 2m + 1)} x^m \times F(\alpha + m, \beta + m; \alpha + \beta - \delta + 2m + 1; x);$$

$$\varphi_m^2 = \varphi_m^2(\alpha, \beta; \delta; x) = \frac{\Gamma(\alpha + \beta - \delta + 2m)}{\Gamma(\alpha + m)\Gamma(\beta + m)} x^{\delta - \alpha - \beta - m} \times F(\delta - \alpha - m, \delta - \beta - m; \delta - \alpha - \beta - 2m + 1; x);$$

$$\varphi_m^3 = \varphi_m^3(\alpha, \beta; \delta; x) = \Gamma(1 - \delta) x^m F(\alpha + m, \beta + m; \delta; 1 - x);$$

(4.2)
$$\varphi_m^4 = \varphi_m^4(\alpha, \beta; \delta; x) = \Gamma(\delta - 1) \frac{\Gamma(\alpha - \delta + m + 1)\Gamma(\beta - \delta + m + 1)}{\Gamma(\alpha + m)\Gamma(\beta + m)} (1 - x)^{1 - \delta} x^m \times F(\alpha - \delta + m + 1, \beta - \delta + m + 1; 2 - \delta; 1 - x);$$

$$\varphi_m^5 = \varphi_m^5(\alpha, \beta; \delta; x) = (-1)^m \Gamma(\alpha - \delta + m + 1) \frac{\Gamma(\beta - \alpha)}{\Gamma(\beta + m)} (-x)^{-\alpha} \times F\left(\alpha + m, \delta - \beta - m; 1 + \alpha - \beta; \frac{1}{x}\right);$$

$$\varphi_m^6 = \varphi_m^6(\alpha, \beta; \delta; x) = \varphi_m^5(\beta, \alpha; \delta; x).$$

Between these branches there are the well-known relations, for instance ([10], §§14.53 and 14.51),

(4.3)
$$\varphi_m^1 = \varphi_m^3 + \varphi_m^4 = \varphi_m^5 + \varphi_m^6 .$$

Any linear combination $\varphi_m = \sum_{i=1}^{6} A_i \varphi_m^i$ with coefficients A_i independent of m and x satisfies the relations

(4.4)
$$x(1 - x)\varphi_m'' = \{(\alpha + \beta + 1)x + \delta - \alpha - \beta - 1\} \varphi_m' + \left\{\alpha\beta + \frac{m}{x}(\alpha + \beta - \delta + m)\right\} \varphi_m ;$$

(4.5)
$$(1 - x)\varphi_m' = m \frac{1 - x}{x} \varphi_m + \frac{(\alpha + m)(\beta + m)}{\alpha + \beta - \delta + 2m + 1} (\varphi_m - \varphi_{m+1});$$

and

$$\frac{\varphi_m}{x} = \left\{ \frac{(\alpha+m)(\alpha-\delta+m+1)+(\beta+m)(\beta-\delta+m+1)}{(\alpha+\beta-\delta+2m-1)(\alpha+\beta-\delta+2m+1)} - \frac{1}{\alpha+\beta-\delta+2m-1} \right\} \varphi_m$$

$$+ \frac{(\alpha+m)(\beta+m)}{(\alpha+\beta-\delta+2m)(\alpha+\beta-\delta+2m+1)}\,\varphi_{m+1}$$

$$+ \frac{(\alpha-\delta+m)(\beta-\delta+m)}{(\alpha+\beta-\delta+2m-1)(\alpha+\beta-\delta+2m)}\,\varphi_{m-1}. \tag{4.6}$$

The first of these formulas follows from the hypergeometric differential equation; the two others can be proved, e.g., by using the hypergeometric series and comparing coefficients of equal powers of x. The last relation corresponds to one of the *relationes inter functiones contiguas*.

From Watson's asymptotic representations of hypergeometric functions we have

$$\lim_{m\to\infty} \frac{\varphi^1_{m+1}}{\varphi^1_m} = \frac{1-(1-x)^{\frac{1}{2}}}{1+(1-x)^{\frac{1}{2}}} \qquad \text{when } \Re(1-x)^{\frac{1}{2}} \geqq 0 \text{ is taken,} \tag{4.7}$$

while

$$\lim_{m\to\infty} \frac{\varphi_{m+1}}{\varphi_m} = \frac{1+(1-x)^{\frac{1}{2}}}{1-(1-x)^{\frac{1}{2}}} \qquad [\Re(1-x)^{\frac{1}{2}} \geqq 0] \tag{4.8}$$

whenever φ_m is not a constant multiple of φ^1_m.

Hence $\sum c_m\varphi^1_m$ and $\sum c_m\varphi_m$ are equiconvergent with the power-series

$$\sum c_m \left(\frac{1-(1-x)^{\frac{1}{2}}}{1+(1-x)^{\frac{1}{2}}}\right)^m \quad \text{and} \quad \sum c_m \left(\frac{1+(1-x)^{\frac{1}{2}}}{1-(1-x)^{\frac{1}{2}}}\right)^m$$

respectively. Let $\lim |\, c_{m+1}/c_m \,| = 1/k$ $(m \to \infty)$. Then the respective domains of convergence are

$$\left|\frac{1-(1-x)^{\frac{1}{2}}}{1+(1-x)^{\frac{1}{2}}}\right| < k \quad \text{and} \quad \left|\frac{1+(1-x)^{\frac{1}{2}}}{1-(1-x)^{\frac{1}{2}}}\right| < k \quad [\Re(1-x)^{\frac{1}{2}} \geqq 0]. \tag{4.9}$$

If $k < 1$, then $\sum c_m\varphi_m$ is divergent, whereas $\sum c_m\varphi^1_m$ is convergent in the domain $|\,1-(1-x)^{\frac{1}{2}}\,| < k\,|\,1+(1-x)^{\frac{1}{2}}\,|$. This domain is bounded by the curve

$$\left|\frac{1-(1-x)^{\frac{1}{2}}}{1+(1-x)^{\frac{1}{2}}}\right| = k \qquad (0 < k < 1), \tag{4.10}$$

which will be shown to be an elliptic limaçon.

Proof. From (4.10),

$$k + \frac{1}{k} = \left|\frac{1-(1-x)^{\frac{1}{2}}}{1+(1-x)^{\frac{1}{2}}}\right| + \left|\frac{1+(1-x)^{\frac{1}{2}}}{1-(1-x)^{\frac{1}{2}}}\right|$$

$$= \frac{|1-(1-x)^{\frac{1}{2}}|^2 + |1+(1-x)^{\frac{1}{2}}|^2}{|x|} = \frac{2}{|x|}\{1+|1-x|\}.$$

Hence (4.10) may be written

$$\frac{1}{2}\left(k+\frac{1}{k}\right)|x| = 1 + |1-x| \qquad \text{or} \qquad \frac{1}{|x|} + \left|1-\frac{1}{x}\right| = \frac{1}{2}\left(k+\frac{1}{k}\right).$$

This last equation shows that $1/x$ lies on an ellipse, i.e., (4.10) is the inverse on $x = 0$ of an ellipse with foci 0, 1 and hence an elliptic limaçon with double focus $x = 0$ and single focus $x = 1$.

If $k > 1$, then $\sum c_m \varphi_m^1$ is convergent in the *whole x-plane* and $\sum \varphi_m c_m$ is convergent outside of the limaçon

$$\left|\frac{1+(1-x)^{\frac{1}{2}}}{1-(1-x)^{\frac{1}{2}}}\right| = k \qquad (1 < k). \tag{4.11}$$

5. Let us assume a solution of Heun's equation in the form

$$y = \sum c_m \varphi_m = \sum_{m=0}^{\infty} c_m(a, h; \alpha, \beta, \gamma, \delta, \epsilon)\varphi_m(\alpha, \beta; \delta; x). \tag{5.1}$$

From (3.1) and (4.4) we easily obtain

$$L[\sum c_m \varphi_m] = \sum c_m[(a-x)\{(\alpha+\beta+1)x + \delta - \alpha - \beta - 1\}\varphi_m'$$

$$+ (a-x)\left\{\alpha\beta + \frac{m}{x}(\alpha+\beta-\delta+m)\right\}\varphi_m + \{\gamma(x-1)(x-a)$$

$$+ \delta x(x-a) + \epsilon x(x-1)\}\varphi_m' + \alpha\beta(x-h)\varphi_m].$$

Using (1.2), this simplifies to

$$\sum c_m\left[\epsilon a(x-1)\varphi_m' + \left\{\alpha\beta(a-h) + m\frac{a-x}{x}(\alpha+\beta-\delta+m)\right\}\varphi_m\right].$$

Employing (4.5) and (1.2) this last expression changes into

$$\sum c_m\left[\left\{\alpha\beta(a-h) - m(\alpha+\beta-\delta+m) + \epsilon a\left(m - \frac{(\alpha+m)(\beta+m)}{\alpha+\beta-\delta+2m+1}\right)\right\}\varphi_m\right.$$

$$\left. + \epsilon a\frac{(\alpha+m)(\beta+m)}{\alpha+\beta-\delta+2m+1}\varphi_{m+1} + m\frac{a}{x}(\gamma+m-1)\varphi_m\right].$$

Finally, by the aid of (4.6), we write the last expression in the form

$$L[\sum c_m\varphi_m] = \sum c_m\{K_{m+1}\varphi_{m+1} + L_m\varphi_m + M_{m-1}\varphi_{m-1}\}, \tag{5.2}$$

where

$$K_{m+1} = a\frac{(\alpha+m)(\beta+m)(\epsilon+m)(\alpha+\beta-\delta+m)}{(\alpha+\beta-\delta+2m)(\alpha+\beta-\delta+2m+1)},$$

$$L_m = am(\gamma+m-1)\left\{\frac{(\alpha+m)(\alpha-\delta+m+1)+(\beta+m)(\beta-\delta+m+1)}{(\alpha+\beta-\delta+2m-1)(\alpha+\beta-\delta+2m+1)}\right.$$

$$\left. -\frac{1}{\alpha+\beta-\delta+2m-1}\right\} - m(\alpha+\beta-\delta+m) - \alpha\beta h$$

$$+ a\frac{\alpha\beta(\gamma+2m)-\epsilon m(\delta-m-1)}{\alpha+\beta-\delta+2m+1} \qquad \text{and}$$

$$M_{m-1} = a\frac{(\alpha-\delta+m)(\beta-\delta+m)m(\gamma+m-1)}{(\alpha+\beta-\delta+2m-1)(\alpha+\beta-\delta+2m)}.$$

(5.3)

Hence (5.1) is a formal solution of Heun's equation if $c_m(a, h; \alpha, \beta, \gamma, \delta, \epsilon)$ is determined by the recurrence formulas

$$L_0c_0 + M_0c_1 = 0,$$

$$K_mc_{m-1} + L_mc_m + M_mc_{m+1} = 0 \qquad (m = 1, 2, 3, \cdots).$$

(5.4)

We shall always put $c_0 = 1$, and c_m shall denote throughout the rest of this paper the coefficients determined by (5.4).

From the recurrence formulas (5.4), the transformation

$$\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha-\delta+m+1)\Gamma(\beta-\delta+m+1)c_m(a, h; \alpha, \beta, \gamma, \delta, \epsilon)$$

$$= \Gamma(\alpha+m)\Gamma(\beta+m)\Gamma(\alpha-\delta+1)\Gamma(\beta-\delta+1)$$

$$\times c_m(a, h_1; \alpha-\delta+1, \beta-\delta+1, \gamma, 2-\delta, \epsilon)$$

(5.5)

immediately follows. h and h_1 are connected by the relation

$$(\alpha-\delta+1)(\beta-\delta+1)h_1 = \alpha\beta h + \gamma(1-\delta)a. \tag{5.6}$$

This transformation is in a way a counterpart to Euler's transformation of the hypergeometric series which reads

$$\varphi_m^1(\alpha, \beta; \delta; x) = \frac{\Gamma(\alpha-\delta+m+1)\Gamma(\beta-\delta+m+1)}{\Gamma(\alpha+m)\Gamma(\beta+m)}$$

$$\times (1-x)^{1-\delta}\varphi_m^1(\alpha-\delta+1, \beta-\delta+1; 2-\delta; x)$$

(5.7)

in the notation adopted in the preceding section.

6. As to the convergence of the series (5.1), two cases must be distinguished. Having $\lim K_m/m^2 = \lim M_m/m^2 = \frac{1}{4}a$ and $\lim L_m/m^2 = \frac{1}{2}a - 1$ $(m \to \infty)$, we infer from a well-known theorem of Poincaré on linear difference equations ([7], cf. also [5], p. 527) that $\lim c_{m+1}/c_m$ $(m \to \infty)$ exists and is equal to one of the roots of the quadratic equation $\frac{1}{4}a\rho^2 + (\frac{1}{2}a - 1)\rho + \frac{1}{4}a = 0$ provided that the moduli of the roots of this equation be distinct, i.e., provided that a is not

a real quantity $\geqq 1$. Henceforward we shall assume $-\pi < \arg(1-a) < \pi$. The two roots of our quadratic are

$$\rho_1 = \frac{1+(1-a)^{\frac{1}{2}}}{1-(1-a)^{\frac{1}{2}}} \quad \text{and} \quad \rho_2 = \rho_1^{-1}. \tag{6.1}$$

Defining the square root uniquely by $\arg(1-a)^{\frac{1}{2}} = \frac{1}{2}\arg(1-a)$, of the two roots ρ_1 has the larger modulus.

In the general case, i.e., if h has arbitrary values, we have $\lim c_{m+1}/c_m = \rho_1$. In the exceptional case, i.e., if h is a root of the transcendental equation ([6], §57)

$$L_0/M_0 - \frac{K_1/M_1}{L_1/M_1 -} \frac{K_2/M_2}{L_2/M_2 -} \frac{K_3/M_3}{L_3/M_3 -} \cdots = 0, \tag{6.2}$$

we have $\lim c_{m+1}/c_m = \rho_2$ $(m \to \infty)$. In the present section we assume the general case and shall deal with the exceptional case later.

In the general case $k = |\rho_1|^{-1}$ (in the notation of section 4) and hence (5.1) is divergent unless φ_m is a multiple of φ_m^1. Hence there is only one solution of (3.1) of type (3.1), namely

$$y_1 = \sum_{m=0}^{\infty} c_m(a, h; \alpha, \beta, \gamma, \delta, \epsilon)\varphi_m^1(\alpha, \beta; \delta; x), \tag{6.3}$$

convergent in the domain

$$\left|\frac{1-(1-x)^{\frac{1}{2}}}{1+(1-x)^{\frac{1}{2}}}\right| < \left|\frac{1-(1-a)^{\frac{1}{2}}}{1+(1-a)^{\frac{1}{2}}}\right| \tag{6.4}$$

bounded by an elliptic limaçon as described in section 4; this limaçon obviously passes through $x = a$. Clearly y_1 belongs to the exponent zero at $x = 0$ and it is $y_1(0) = 1$. Hence

$$\frac{\Gamma(\alpha+\beta-\delta+1)}{\Gamma(\alpha-\delta+1)\Gamma(\beta-\delta+1)}\, y_1 = F(a, h; \alpha, \beta, \gamma, \delta, \epsilon; x) \tag{6.5}$$

is the fundamental solution studied by Heun of equation (3.1). Heun proved that any solution of (3.1) can be expressed in terms of his function (6.5). Hence any solution may be expressed in series of the type (5.1).

(3.1) has 192 solutions of type (6.5) ([10], Chapter 23, Example 10), 48 of which have been given by Heun. The rest may be obtained by the transformations of Heun's series we are going to develop now.

By the transformations (5.5) and (5.7) we have

$$\begin{aligned}\Gamma(\alpha)\Gamma(\beta)y_1 &= \sum \Gamma(\alpha)\Gamma(\beta)c_m\varphi_m^1 \\ &= (1-x)^{1-\delta}\Gamma(\alpha-\delta+1)\Gamma(\beta-\delta+1)\sum c_m(a, h_1; \\ &\quad \alpha-\delta+1, \beta-\delta+1, \gamma, 2-\delta, \epsilon)\varphi_m^1(\alpha-\delta+1, \beta-\delta+1; 2-\delta; x)\end{aligned}$$

and consequently

$$\begin{aligned}&F(a, h; \alpha, \beta, \gamma, \delta, \epsilon; x) \\ &\qquad = (1-x)^{1-\delta}F(a, h_1; \alpha-\delta+1, \beta-\delta+1, \gamma, 2-\delta, \epsilon; x),\end{aligned} \tag{6.6}$$

where h and h_1 are connected by (5.6).

Also it is easily seen that

$$F(a, h; \alpha, \beta, \gamma, \delta, \epsilon; x) = F\left(\frac{1}{a}, \frac{hF}{a}; \alpha, \beta, \gamma, \epsilon, \delta; \frac{x}{a}\right) \tag{6.7}$$

and applying (6.6) to the right hand side of (6.7), we obtain the third transformation

$$\begin{aligned} &F(a, h; \alpha, \beta, \gamma, \delta, \epsilon; x) \\ &= \left(1 - \frac{x}{a}\right)^{1-\epsilon} F\left(\frac{1}{a}, \frac{h_2}{a}; \alpha - \epsilon + 1, \beta - \epsilon + 1, \gamma, 2 - \epsilon, \delta; \frac{x}{a}\right) \\ &= \left(1 - \frac{x}{a}\right)^{1-\epsilon} F(a, h_2; \alpha - \epsilon + 1, \beta - \epsilon + 1, \gamma, \delta, 2 - \epsilon; x), \end{aligned} \tag{6.8}$$

where h and h_2 are connected by the relation

$$(\alpha - \epsilon + 1)(\beta - \epsilon + 1)h_2 = \alpha\beta h + \gamma(1 - \epsilon). \tag{6.9}$$

The series of hypergeometric functions corresponding to the right hand sides of (6.7) and (6.8) are convergent in a domain of the x-plane bounded by an elliptic limaçon with double focus 0 and single focus a and passing through $x = 1$.

7. For the rest of this paper let us suppose that h is a root of (6.2). Then $k = |\rho_2|^{-1} > 1$ and consequently the series (6.3) is convergent in the whole x-plane and represents a solution regular at $x = 0$ and $x = a$. There is a branch-cut along the real axis extending from $x = 1$ to $x = +\infty$. The solutions

$$y_i = \sum_{m=0}^{\infty} c_m(a, h; \alpha, \beta, \gamma, \delta, \epsilon)\varphi_m^i(\alpha, \beta; \delta; x) \quad (i = 2, \cdots, 6) \tag{7.1}$$

are convergent outside of the elliptic limaçon with foci $x = 0$ and 1 and passing through $x = a$. Hence all solutions are convergent in domains of which $x = 1$ and $x = \infty$ are inner points. Thus the substitutions of the monodromic group are known as far as they are related to these two singularities. At $x = 1$ and $x = \infty$ the monodromic group of (1.1) is isomorphic to the group of

$$P\left\{\begin{matrix} 0 & 1 & \infty \\ 0 & 0 & \alpha \\ \delta - \alpha - \beta & 1 - \delta & \beta \end{matrix} \quad x\right\}.$$

y_3 and y_4 are the fundamental solutions at $x = 1$; y_5 and y_6 are the fundamental solutions belonging to $x = \infty$. From (4.3) we obtain

$$y_1 = y_3 + y_4 = y_5 + y_6, \tag{7.2}$$

and all the other relations from the theory of the hypergeometric function.

In this exceptional case y_1 may be developed into a series of Jacobi polynomials (Svartholm, loc. cit.)

$$y_1 = \sum_{m=0}^{\infty} A_m F\left(-m, \gamma + \epsilon + m - 1; \gamma; \frac{x}{a}\right) \tag{7.3}$$

convergent in a domain of the x-plane bounded by an ellipse with foci $x = 0$ and $x = a$ and passing through $x = 1$. No other solution of (3.1) can be developed into a similar series in general. I may mention without proof that the expansions (7.3) and (6.3) are connected with each other by Lambe-Ward's integral equation. More precisely, the integral equation transforms (7.3) into (6.3), just as Whittaker's integral equation for Lamé functions transforms the Fourier-Jacobi expansion of these functions into their expansion in series of Legendre functions. In fact, Lambe-Ward's integral equation is equivalent with the co-existence and identity of (7.3) and (6.3). It seems that Lambe-Ward's integral equation being restricted to the exceptional case is intimately connected with there being in the general case (i.e., with arbitrary values of h) no expansion of type (7.3).

It is hardly necessary to mention that beside the exceptional case dealt with in this section, there are other exceptional cases in which Heun's equation has a solution regular at two singularities other than 0 and a of this equation. All these cases are obtainable by linear transformations from the one dealt with here and so need not be considered separately.

Also, there are certain exceptional cases in which there is a solution regular at three singularities. If we take $\alpha = -M$, $M = 0, 1, 2, \cdots$, and h to be a root of (6.2) (which is an algebraic equation of degree M in this case), we have $K_{M+1} = 0$ and $c_{M+1} = 0$. Hence also $c_{M+2} = c_{M+3} = \cdots = 0$ and all series (5.1) terminate. y_1 is in this case a polynomial and identical with y_5 (Heun polynomial); y_6 is Heun's function of the second kind. The further particulars are so similar to the corresponding results on Lamé polynomials that it is not necessary to go into details.

References

1. A. Erdélyi, *Expansion of Lamé functions into series of Legendre functions*, in press.
2. K. Heun, *Zur Theorie der Riemann'schen Funktionen zweiter Ordnung mit vier Verzweigungspunkten*, Math. Annalen, vol. 33(1889), pp. 161–179.
3. E. L. Ince, *Ordinary Differential Equations*, London, 1927.
4. C. G. Lambe and D. R. Ward, *Some differential equations and associated integral equations*, Quart. J. Math. (Oxford), vol. 5(1934), pp. 81–97.
5. L. M. Milne-Thomson, *The Calculus of Finite Differences*, London, 1933.
6. O. Perron, *Die Lehre von den Kettenbrüchen*, Leipzig, 1913.
7. H. Poincaré, *Sur les équations linéaires aux différentielles ordinaires et aux différences finies*, Amer. J. Math., vol. 7(1885), pp. 203–258.
8. N. Svartholm, *Die Lösung der Fuchsschen Differentialgleichung zweiter Ordnung durch hypergeometrische Polynome*, Math. Annalen, vol. 116(1939), pp. 413–421.
9. G. N. Watson, *Asymptotic expansions of hypergeometric functions*, Trans. Cambridge Phil. Soc., vol. 22(1918), pp. 277–308.
10. E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, Cambridge, 1927.

University of Edinburgh.

A GENERALIZATION OF THE EUCLIDEAN ALGORITHM TO SEVERAL DIMENSIONS

By Barkley Rosser

Summary. The Euclidean algorithm is generalized to two, three, and four dimensions. The generalized algorithm is applied to the solution of the following problems.

Given a positive definite quadratic form, find integer values of the variables, not all zero, which make the value of the form a minimum.

Given n linear forms in n variables with determinant $\Delta \neq 0$, find integer values of the variables, not all zero, such that each linear form $\leqq |\Delta|^{1/n}$ in absolute value.

Given n real numbers, $x_1, \cdots, x_n$, not all rational, find as many sets, $a, a_1, a_2, \cdots, a_n$, of integers as desired such that simultaneously

$$|ax_i - a_i| \leqq a^{-1/n} \qquad (i = 1, 2, \cdots, n).$$

Given

$$L = \sum_{i=1}^{n} a_i x_i,$$

the a_i's being coprime integers, find a general solution, in integers, of $L = k_1$, namely

$$x_i = \sum_{j=1}^{n} b_{ij} k_j \qquad (i = 1, \cdots, n)$$

(where the b's are fixed integers, k_1 is the same integer that occurs in $L = k_1$, and the other k's are arbitrary integers) such that $\sum (b_{ij})^2$ shall be a minimum.

Given a hypersphere, σ, with center at the origin and radius $\geqq 1$, and

$$L = \sum_{i=1}^{n} u_i x_i,$$

the u's being real numbers, find a lattice point distinct from the origin within (or on) σ and as close to the hyperplane $L = 0$ as possible.

Given two symmetric positive definite matrices, A and B, with real components, find whether there is a matrix P with integral components and determinant ± 1 such that $B = P^T A P$, and if so, to find all such P's.

We open the paper with some preliminary conventions regarding terminology.

We shall use lower case italics from a to t inclusive for rational integers, and from u to z inclusive for real numbers. We shall use upper case italics from A to Q inclusive for square matrices with real elements, and from R to Z in-

Received August 9, 1941; presented to the American Mathematical Society, May 2, 1941. A summary of this paper appeared under the same title in Proceedings of the National Academy of Sciences, vol. 27(1941), pp. 309–311.

clusive for vectors with real components. A^T denotes the transpose of A. Vectors will be thought of as matrices of one column when convenient, so that the inner product of U and V can be written as the matrix product U^TV. Hence the length of U (which we will denote by $L(U)$) is $(U^TU)^{1/2}$, and the cosine of the angle between U and V is

$$\frac{U^TV}{L(U)L(V)}.$$

At other times it will be convenient to think of a vector as a point in n-dimensional space, namely, as the point whose coordinates are the components of the vector. Thus we can speak of a limit point of a set of vectors. Any finite sum of $X_1, \cdots, X_s$ with integer coefficients, say $\sum a_iX_i$, will be called an I. L. C. (integral linear combination) of the X's.

As our first step in generalizing the Euclidean algorithm, we will generalize one of the problems which this algorithm solves, namely, the problem of finding a greatest common factor of two integers. This generalization will be much facilitated if we choose an appropriate definition of the G. C. F. of two integers. Among the many possible, we choose the following.

f is a G. C. F. of a and b if and only if:

1. there are integers m and n such that $f = ma + nb$,
2. there are integers d_1 and d_2 such that $a = d_1f$ and $b = d_2f$.

As it stands, this is a definition of an integer f being a G. C. F. of two integers a and b. If we replace a, b, and f by x, y, and z, then we have a definition of a real number z being a G. C. F. of two real numbers x and y. However, two real numbers, x and y, can be identified with two collinear vectors of lengths x and y, and conversely. Hence we consider replacing a, b, and f by X, Y, and Z, and then we have a definition of a vector Z being a G. C. F. of two vectors X and Y. By condition 2, X and Y would have to be collinear.

It is clear that an arbitrary pair of collinear vectors need not have a G. C. F. and that a necessary and sufficient condition that they should is that they be commensurable (that is, have commensurable lengths). Here again, generalization is facilitated by an appropriate choice of a definition of "commensurable". In fact, there is available a definition in which the generalization is exactly as simple as the special case. We present the generalization.

DEFINITION 1. A set of vectors $V_1, V_2, \cdots, V_n$ is said to be commensurable if and only if the set of I. L. C.'s of the V's has no limit point.

Here and later, n is not to be construed as having any connection with the dimensionality of the space in which the V's lie.

We now present the generalized definition of G. C. F.

DEFINITION 2. A set of vectors $U_1, \cdots, U_m$ is a G. C. F. of a set of vectors $V_1, \cdots, V_n$ if and only if:

1. each U is an I. L. C. of the V's,
2. each V is an I. L. C. of the U's,
3. the U's are linearly independent.

Note that conditions 1 and 2 of Definition 2 are strict generalizations of conditions 1 and 2 of the definition in the linear case, and condition 3 corresponds to the fact that in the linear case a G. C. F. consists of a single vector.

We now undertake to generalize the Euclidean algorithm to a form which can be used to find a G. C. F. of a set of commensurable vectors. There are numerous minor variations of the Euclidean algorithm, and it will be helpful to start with an appropriate variation. We shall choose the variation known as the least remainder algorithm, and for reference shall explain how it would be applied to the problem of finding a G. C. F. of two collinear vectors.

Let U_1 and V_1 be two commensurable, collinear vectors, not both zero. If $U_1 = 0$, then V_1 is a G. C. F. of U_1 and V_1. So assume $U_1 \neq 0$. Since U_1 and V_1 are collinear, we can find a real x such that $V_1 = xU_1$. Choose m a nearest integer to x (clearly m is unique unless x is halfway between two integers). Put $U_2 = V_1 - mU_1$, $V_2 = U_1$. Then $U_2 = V_1 - xU_1 + (x - m)U_1 = 0 + (x - m)U_1$. Since m is a nearest integer to x, $|x - m| \leqq \frac{1}{2}$. So $L(U_2) \leqq \frac{1}{2}L(U_1)$. If $U_2 \neq 0$, we can repeat the process, getting $U_3 = V_2 - nU_2$, $V_3 = U_2$, $L(U_3) \leqq \frac{1}{2}L(U_2)$. Moreover, we can continue to repeat the process as long as $U_n \neq 0$, and we will continue to have $L(U_{n+1}) \leqq \frac{1}{2}L(U_n)$. Also each of U_{n+1} and V_{n+1} will be an I. L. C. of U_n and V_n. Note that conversely each of U_n and V_n will be an I. L. C. of U_{n+1} and V_{n+1}. From this it follows that, for every n, each of U_n and V_n is an I. L. C. of U_1 and V_1, and each of U_1 and V_1 is an I. L. C. of U_n and V_n. Suppose there is never an n such that $U_n = 0$. Then we would have an infinite succession of U_n's, each no more than half as long as the preceding, and each an I. L. C. of U_1 and V_1. This would mean that the origin is a limit point of the set of I. L. C.'s of U_1 and V_1, which would contradict our assumption that U_1 and V_1 are commensurable. So there must be an n such that $U_n = 0$. Then V_n is an I. L. C. of U_1 and V_1, and conversely each of U_1 and V_1 is an I. L. C. of V_n. So V_n is a G. C. F. of U_1 and V_1.

We now generalize to two dimensions. Let U_1, V_1, W_1 be three commensurable, coplanar vectors, not all zero. If two are zero, we are reduced to the case of two commensurable collinear vectors, which we have already discussed. So we assume that not more than one of U_1, V_1, and W_1 is zero. If U_1, V_1, W_1 are collinear, then find a G. C. F., X, of U_1 and V_1. As this is an I. L. C. of U_1 and V_1, any I. L. C. of X and W_1 is an I. L. C. of U_1, V_1, and W_1. So X and W_1 are commensurable. Also, X and W_1 are collinear so that we can find a G. C. F., Y, of X and W_1. Y is clearly a G. C. F. of U_1, V_1, and W_1. Now we turn to the case where U_1, V_1, and W_1 are not collinear. There is clearly no loss of generality in assuming $L(U_1) \leqq L(V_1) \leqq L(W_1)$.

Case 1. U_1 and V_1 are collinear. Find a G. C. F., X, of U_1 and V_1. Then X and W_1 are independent. Hence X and W_1 constitute a G. C. F. of U_1, V_1, and W_1.

Case 2. U_1 and V_1 are not collinear. Then, since U_1, V_1, W_1 are coplanar,

W_1 must be a linear combination of U_1 and V_1, say $W_1 = xU_1 + yV_1$. Choose m and n integers nearest to x and y respectively, and put $W = W_1 - mU_1 - nV_1$. Then $W = W_1 - xU_1 - yV_1 + (x - m)U_1 + (y - n)V_1 = (x - m)U_1 + (y - n)V_1$. Since U_1 and V_1 are not collinear,

$$\begin{aligned} L((x - m)U_1 + (y - n)V_1) &< L((x - m)U_1) + L((y - n)V_1) \\ &\leqq \tfrac{1}{2}L(U_1) + \tfrac{1}{2}L(V_1) \\ &\leqq \tfrac{1}{2}L(V_1) + \tfrac{1}{2}L(V_1). \end{aligned}$$

So $L(W) < L(V_1)$. Now take U_2 to be the shorter of U_1 and W (if $L(U_1) = L(W)$, take $U_2 = U_1$), V_2 to be the other of U_1 and W, and W_2 to be V_1. Then $L(U_2) + L(V_2) + L(W_2) < L(U_1) + L(V_1) + L(W_1)$. Also $L(U_2) \leqq L(V_2) \leqq L(W_2)$. So if U_2 and V_2 are not collinear, we can repeat the process. In fact, we can continue to repeat the process as long as U_n and V_n are not collinear. Note that we will have $L(U_{n+1}) + L(V_{n+1}) + L(W_{n+1}) < L(U_n) + L(V_n) + L(W_n)$ at each step. Also, each of U_n, V_n, W_n is an I. L. C. of U_1, V_1, W_1, and conversely. Now suppose that we always have U_n and V_n collinear. Then we get an infinite succession of U_n's, V_n's, and W_n's. Among all these there must be an infinite number of distinct points, since otherwise we could not have $L(U_{n+1}) + L(V_{n+1}) + L(W_{n+1}) < L(U_n) + L(V_n) + L(W_n)$ for all n. Also each U_n, V_n, or W_n must have length $\leqq L(U_n) + L(V_n) + L(W_n) \leqq L(U_1) + L(V_1) + L(W_1)$. Hence, we have an infinite number of distinct U_n's, V_n's, and W_n's lying within a circle of radius $L(U_1) + L(V_1) + L(W_1)$ with center at the origin. So they must have a limit point. As they are all I. L. C.'s of U_1, V_1, and W_1, this contradicts our assumption that U_1, V_1, and W_1 are commensurable. So for some n, U_n and V_n are collinear. Then we can apply Case 1 and find a G. C. F. of U_n, V_n, W_n, and this will be a G. C. F. of U_1, V_1, W_1.

A short digression is in order here. Just as the Euclidean algorithm can be applied to two incommensurable real numbers, yielding the continued fraction algorithm for approximating an irrational, so the algorithm just explained can be applied to three incommensurable vectors, and yields an algorithm for the simultaneous approximation of two irrationals. For instance, suppose we put

$$U_1 = (1, 0), \qquad V_1 = (0, 1), \qquad W_1 = (\sqrt{2}, \sqrt{3}).$$

Then $W_1 = 1.41U_1 + 1.73V_1$. So we take

$$U_2 = W_1 - U_1 - 2V_1 = (\sqrt{2} - 1, \sqrt{3} - 2), \quad V_2 = U_1, \quad W_2 = V_1.$$

Then $W_2 = -3.73U_2 + 1.55V_2$. So we take

$$U_3 = W_2 + 4U_2 - 2V_2 = (4\sqrt{2} - 6, 4\sqrt{3} - 7), \quad V_3 = U_2, \quad W_3 = V_2.$$

Then $W_3 = -2.20U_3 + .59V_3$. So we take

$$U_4 = W_3 + 2U_3 - V_3 = (7\sqrt{2} - 10, 7\sqrt{3} - 12), \quad V_4 = U_3, \quad W_4 = V_3.$$

Then $W_4 = -2.44U_4 - .49V_4$. So we take

$$U_5 = U_4, \quad V_5 = W_4 + 2U_4 = (15\sqrt{2} - 21, 15\sqrt{3} - 26), \quad W_5 = V_4.$$

Then $W_5 = -.89U_5 - 2.03V_5$. So we take

$$U_6 = W_5 + U_5 + 2V_5 = (41\sqrt{2} - 58, 41\sqrt{3} - 71), \quad V_6 = U_5, \quad W_6 = V_5.$$

And so on. Note that $L(U_2) > L(U_3) > L(U_4) > L(U_6)$. As the length of a vector is an upper bound on its components, there is a sense in which we can say that the vectors U_2, U_3, U_4, U_6 have shortening components. As the components have the form $(a\sqrt{2} - b, a\sqrt{3} - c)$, we have here a scheme for choosing integers a, b, and c which make $a\sqrt{2} - b$ and $a\sqrt{3} - c$ simultaneously small. Dividing through by a, we see that we are simultaneously approximating $\sqrt{2}$ and $\sqrt{3}$ by means of fractions with the same denominator. For the instance at hand, we can say even more. U_2, U_3, U_4, V_5, and U_6 all furnish values of a, b, and c which satisfy

$$\left|\sqrt{2} - \frac{b}{a}\right| < \frac{1}{a\sqrt{a}}, \qquad \left|\sqrt{3} - \frac{c}{a}\right| < \frac{1}{a\sqrt{a}}.$$

Naturally this raises a number of questions, such as the following.

Will the algorithm continue to give this degree of approximation indefinitely?

Will the algorithm give a similar degree of approximation for other pairs of irrationals?

What sorts of irrationals, if any, will cause the algorithm to repeat indefinitely?

We will not attempt to answer these questions in this paper, but will proceed with our generalization.

The proof that $L(W) < L(V_1)$ which we give in the two dimensional case will obviously not generalize to any more dimensions. A possible way out of this difficulty is suggested by the observation that the W which we have chosen is not always the shortest possible W. For example, looking back at our numerical illustration, $W_4 = -2.44U_4 - .49V_4$, so that according to specifications W should be $W_4 + 2U_4$. However, both $W_4 + 2U_4 + V_4$ and $W_4 + 3U_4$ are shorter. In fact, each of these is shorter than U_4, whereas $W_4 + 2V_4$ is only shorter than V_4. Clearly it would improve the algorithm if we take W to be the shortest vector of the form $W_i - mU_i - nV_i$ at each step. Naturally this raises the question of how to determine m and n so as to minimize the length of $W_i - mV_i - nU_i$. This question will be answered by Construction 2 below, for which we now prepare the way.

CONSTRUCTION 1. Given the independent vectors $V_1, V_2, \cdots, V_n$, the arbitrary vector U, and the positive constant z, find all vectors, $W = U - \sum m_i V_i$, such that $L(W) \leqq z$.

We shall give the construction by induction on n. That is, we shall first describe the construction for $n = 1$. Then assuming that the construction can be carried out for n, we shall show how to carry it out for $n + 1$.

Let $n = 1$. The condition that V_1 constitutes an independent set of vectors

implies $V_1 \neq 0$, and hence $L(V_1) \neq 0$. We wish all values of m such that $L(U - mV_1) \leqq z$, that is, all m's such that $(U - mV_1)^T(U - mV_1) \leqq z^2$ or $U^TU - 2mU^TV_1 + m^2V_1^TV_1 \leqq z^2$. Plot the parabola $y = U^TU - 2xU^TV_1 + x^2V_1^TV_1$. Then the values of x for which $y \leqq z^2$ form a finite closed interval (or perhaps a single point, or perhaps the null set), and we choose for m the integer values of x lying in that interval.

Suppose we can perform the construction for n. Let $V_1, \cdots, V_n, V_{n+1}$ be independent. Let U_2 and X_2 be the the projections of U and V_{n+1} on the linear manifold of $V_1, \cdots, V_n$, and put $U_1 = U - U_2$ and $X_1 = V_{n+1} - X_2$. Then U_1 and X_1 are orthogonal to the linear manifold of $V_1, \cdots, V_n$. If $W = U - \sum m_iV_i$, put $W_1 = U_1 - m_{n+1}X_1$ and $W_2 = U_2 - m_{n+1}X_2 - m_1V_1 - \cdots - m_nV_n$. Then $W = W_1 + W_2$, and W_1 and W_2 are orthogonal. Hence $[L(W)]^2 = [L(W_1)]^2 + [L(W_2)]^2$. By the construction for $n = 1$, find all values of m_{n+1} such that $L(W_1) = L(U_1 - m_{n+1}X_1) \leqq z$. For each such m_{n+1}, use the construction for n to find all sets of values for $m_1, \cdots, m_n$ which make $L(W_2) = L((U_2 - m_{n+1}X_2) - m_1V_1 - \cdots - m_nV_n) \leqq (z^2 - [L(W_1)]^2)^{1/2}$.

THEOREM 1. *If $V_1, \cdots, V_n$ are independent, then there are only a finite number of vectors, $W = U - \sum m_iV_i$, for which $L(W) \leqq z$.*

Proof. Observe that Construction 1, which yields all such W's, yields only a finite number.

THEOREM 2. *If $V_1, \cdots, V_n$ are independent, then there is a shortest non-zero vector, $W = U - \sum m_iV_i$.*

Proof. U and $U - V_1$ are not both zero. Hence there is a non-zero W. Choose a non-zero W. By Theorem 1, there are only a finite number of shorter W's, and so there is a shortest non-zero W.

Note that it is not claimed that there is a unique shortest non-zero W. We shall see later that it is quite possible to have several shortest non-zero W's (all the same length, of course).

CONSTRUCTION 2. Given the independent vectors $V_1, \cdots, V_n$, and an arbitrary vector U, find a shortest non-zero vector, $W = U - \sum m_iV_i$.

The construction is indicated in the proof of Theorem 2. We will, however, make some suggestions for abridging the computations. Note that the method consists of finding a non-zero W, and then finding all shorter W's. The shorter our original W is, the fewer shorter W's there are, and so the less labor to find all of them. This suggests that as shorter W's are found, their lengths be used as new upper bounds instead of the lengths of the earlier and longer W's. Also when a non-zero W has been found, and we start in to find all shorter W's, we should try the shortest W_1 first (see the latter part of the instructions for Construction 1). Recall that $W_1 = U_1 - m_{n+1}X_1$, and that the (m_{n+1})'s are determined by the construction for $n = 1$, which is such that it is easy to choose the W_1's in order of increasing length.

A word of caution is desirable. One might suppose that if $U = xV_1 + yV_2$,

then the minimum length for $U - mV_1 - nV_2$ would be attained with an m and n which are close to x and y respectively. Such is not always the case. Let $U = (2, 2)$, $V_1 = (47, 7)$, $V_2 = (13, 2)$. Then

$$U = -7\tfrac{1}{3}V_1 + 26\tfrac{2}{3}V_2 .$$

Nevertheless the shortest vector of the form $U - mV_1 - nV_2$ is obtained by putting $m = -11$, $n = 40$. For more dimensions, the situation is obviously no better.

We now return to the question of generalizing the algorithm which we elucidated for the case of three commensurable coplanar vectors. It was suggested that if we modify that algorithm by taking W to be the shortest vector of the form $W_1 - mU_1 - nV_1$, the resulting algorithm can be generalized to more dimensions without further modification. This is the case, and the next two theorems supply the information needed to show that the generalization is effective for three and four dimensions.

THEOREM 3. *Let U, V, W, X be dependent vectors. Let $L(U) \leqq L(V) \leqq L(W)$. Let U, V, W be independent. Let Y be a shortest vector of the form $X - mU - nV - pW$. Then $L(Y) < L(W)$.*

The proof is just like the proof of the next theorem.

THEOREM 4. *Let U, V, W, X, Y be dependent vectors. Let $L(U) \leqq L(V) \leqq L(W) \leqq L(X)$. Let U, V, W, X be independent. Let Z be a shortest vector of the form $Y - mU - nV - pW - qX$. Then $L(Z) < L(X)$ except in the special case where U, V, W, X are mutually orthogonal and all the same length, and $Z = \pm\frac{1}{2}U \pm \frac{1}{2}V \pm \frac{1}{2}W \pm \frac{1}{2}X$.*

Note. In the exceptional case, U, V, W, and Z would constitute a G. C. F. of U, V, W, X, Y. Hence, the exceptional case merely provides an additional way in which the algorithm may terminate in the case of four dimensions.

Proof. Choose a coordinate system in which we have

$$U = (u_1 , 0, 0, 0, 0, \cdots , 0),$$

$$V = (v_1 , v_2 , 0, 0, 0, \cdots , 0),$$

$$W = (w_1 , w_2 , w_3 ; 0, 0, \cdots , 0),$$

$$X = (x_1 , x_2 , x_3 , x_4 , 0, \cdots , 0),$$

$$Y = (y_1 , y_2 , y_3 , y_4 , 0, \cdots , 0).$$

Then $|u_1| = L(U)$, $|v_2| \leqq L(V)$, $|w_3| \leqq L(W)$, and $|x_4| \leqq L(X)$. Moreover, $|v_2| = L(V)$ if and only if $v_1 = 0$, $|w_3| = L(W)$ if and only if $w_1 = w_2 = 0$, and $|x_4| = L(X)$ if and only if $x_1 = x_2 = x_3 = 0$. Now choose q so that the fourth component of $Y - qX$ is numerically $\leqq |x_4/2|$. Then choose p so that the third component of $Y - pW - qX$ is numerically $\leqq |w_3/2|$. Then choose n so that the second component of $Y - nV - pW - qX$ is numerically $\leqq |v_2/2|$. Finally choose m so that the first component of $Y - mU - nV - pW - qX$

is numerically $\leqq |u_1/2|$. Denote $Y - mU - nV - pW - qX$ by R. Then $L(R)$ is the square root of the sum of the squares of the components of R, and hence

$$L(R) \leqq \left(\frac{u_1^2}{4} + \frac{v_2^2}{4} + \frac{w_3^2}{4} + \frac{x_4^2}{4}\right)^{1/2}.$$

So

$$L(R) \leqq \tfrac{1}{2}([L(U)]^2 + [L(V)]^2 + [L(W)]^2 + [L(X)]^2)^{1/2}.$$

However, $L(U) \leqq L(V) \leqq L(W) \leqq L(X)$. So $L(R) \leqq L(X)$. Furthermore, Z is a shortest vector of the form $Y - mU - nV - pW - qX$. So $L(Z) \leqq L(R)$. So $L(Z) \leqq L(X)$. Now we need to consider the circumstances under which we can have $L(Z) = L(X)$. Clearly this can happen only when $L(R) = L(X)$. This latter can happen only when $L(U) = L(V) = L(W) = L(X) = |u_1| = |v_2| = |w_3| = |x_4|$. So we must have $v_1 = w_1 = w_2 = x_1 = x_2 = x_3 = 0$. So U, V, W, X are mutually perpendicular and all the same length. Moreover, the components of R must each have absolute value equal to $\frac{1}{2}L(U)$. Hence $R = \pm\frac{1}{2}U \pm \frac{1}{2}V \pm \frac{1}{2}W \pm \frac{1}{2}X$. Under these circumstances, any shortest vector (and hence Z) must have the same form as R.

The generalization of Theorem 5 to five dimensions is false, as one can see by putting

$$U = (1, 0, 0, 0, 0),$$

$$V = (0, 1, 0, 0, 0),$$

$$W = (0, 0, 1, 0, 0),$$

$$X = (0, 0, 0, 1, 0),$$

$$Y = (0, 0, 0, 0, 1),$$

$$Z = (\tfrac{5}{11}, \tfrac{5}{11}, \tfrac{5}{11}, \tfrac{5}{11}, \tfrac{5}{11}).$$

Clearly Z is the shortest vector of the form $Z - mU - nV - pW - qX - rY$. However Y is not the shortest vector of the form $Y - mU - nV - pW - qX - rZ$. In fact, if we put $R = Y + U + V + W + X - 2Z$, then $L(R) < L(Y)$. Moreover U, V, W, X, R constitute a G. C. F. for U, V, W, X, Y, Z. This suggests a further modification of the algorithm for five or more dimensions. Instead of invariably trying to shorten the longest vector by subtracting an I. L. C. of the other vectors, try in turn to shorten each vector in this fashion. Whether this modification will restore the effectiveness of the algorithm for five dimensions, I do not know. That even this modification can fail for a large number of dimensions is shown by the following example. For $i = 1, 2, \cdots, 26$, let V_i be the vector with twenty-six components, of which the i-th is unity and the rest are zero. Let V_{27} be the vector with twenty-six components, all equal to 2/5. Clearly, no one of the twenty-seven V's can

be shortened by subtracting an I. L. C. of the other vectors. Nevertheless, they have a G. C. F., namely, $V_1, \cdots, V_{25}$ and $\frac{1}{2}V_{27}$.

In spite of the failure of the algorithm for a large number of dimensions, we have the following theorem.

THEOREM 5. *Any n commensurable vectors, not all zero, have a* G. C. F.

The proof is by strong induction on n. When $n = 1$, matters are simple. Assume the theorem true for n or less vectors and let us have $n + 1$ vectors, $V_1, \cdots, V_n, V_{n+1}$. If they are independent, they themselves constitute their G. C. F. Now let one of them, V_{n+1} (say), be dependent on the rest. If $V_1, \cdots, V_n$ are dependent, they have a G. C. F., $U_1, \cdots, U_s$ with $s < n$. Also any I. L. C. of V_{n+1} and the U's is an I. L. C. of the V's. So V_{n+1} and the U's are commensurable. As they are at most n in number, they have a G. C. F., which is a G. C. F. of the V's. So we have remaining the case where $V_1, \cdots, V_n$ are independent, and V_{n+1} is dependent on $V_1, \cdots, V_n$. Then choose coordinates in which the V's have components as indicated:

$$V_1 = (v_{11}, 0, 0, 0, \cdots, 0, 0, \cdots, 0),$$

$$V_2 = (v_{21}, v_{22}, 0, 0, \cdots, 0, 0, \cdots, 0),$$

$$V_3 = (v_{31}, v_{32}, v_{33}, 0, \cdots, 0, 0, \cdots, 0),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$V_n = (v_{n1}, v_{n2}, v_{n3}, v_{n4}, \cdots, v_{nn}, 0, \cdots, 0),$$

$$V_{n+1} = (w_1, w_2, w_3, w_4, \cdots, w_n, 0, \cdots, 0).$$

Then v_{nn} and w_n are commensurable. For, suppose they are not. Choose sets of integers m_i, n_i such that $m_i v_{nn} + n_i w_n$ converge to zero. Consider $X_i = m_i V_n + n_i V_{n+1} + a_{i1}V_1 + a_{i2}V_2 + a_{i,n-1}V_{n-1}$. By properly choosing first $a_{i,n-1}$, then $a_{i,n-2}$, $\cdots$, and finally a_{i1}, we can make the j-th component of X_i less in absolute value than v_{jj} for $1 \leqq j \leqq n - 1$. As the n-th components of the X's are all different, we have an infinite number of X's in a bounded region, and so the X's have a limit point, contradicting the commensurability of the V's. Hence there is a positive number z and integers a, b, c, and d such that $av_{nn} + bw_n = z$, $cz = v_{nn}$, $dz = w_n$. Then $ac + bd = 1$ as we can see by substituting the expressions for v_{nn} and w_n into the expression for z. Now define $Y_1 = dV_n - cV_{n+1}$, $Y_2 = aV_n + bV_{n+1}$. Then $bY_1 + cY_2 = V_n$ and $dY_2 - aY_1 = V_{n+1}$. So any G. C. F. of $V_1, \cdots, V_{n-1}, Y_1, Y_2$ is a G. C. F. of the V's, and conversely. Now the n vectors $V_1, \cdots, V_{n-1}, Y_1$ have a G. C. F., $U_1, \cdots, U_s$. Since the n-th components of $V_1, \cdots, V_{n-1}, Y_1$ are all zero, the same is true of the U's. Hence Y_2 and the U's form an independent set of vectors, and hence constitute a G. C. F. of the V's.

Note that if one follows the proof of Theorem 5, one can describe a constructive procedure for finding a G. C. F. of a set of commensurable vectors. However, it is not one that can be recommended for ease of application, and a person would do well to try modifications of the algorithms described in this paper

(especially the one which is still to be described) as long as they are effective, and save the procedure based on the proof of Theorem 5 as a last resort.

For completeness, we digress long enough to indicate that the converse of Theorem 5 is true. Suppose $V_1, \cdots, V_n$ have a G. C. F. $U_1, \cdots, U_s$, but the V's are incommensurable. As every I. L. C. of the V's is an I. L. C. of the U's, the U's are incommensurable. So a sequence of I. L. C.'s of the U's, $X_1, X_2, \cdots$, approaches a limit X. So if one chooses a positive ϵ, there are an infinite number of vectors $X - X_i$ with $L(X - X_i) < \epsilon$. This contradicts Theorem 1.

We pause for orientation. In one dimension a G. C. F. of two collinear commensurable non-zero vectors is a shortest I. L. C. of the two vectors, and conversely. However, in more dimensions, a G. C. F. of a set of vectors has not the least connection with a shortest I. L. C. of the vectors. (To see this, note that if U_1, U_2 constitute a G. C. F. of $V_1, \cdots, V_n$, then $aU_1 + bU_2, cU_1 + dU_2$ also constitute a G. C. F. if and only if $ad - bc = 1$.) In the remainder of the paper we shall be concerned with another generalization of the Euclidean algorithm, which will solve the problem of finding a shortest non-zero I. L.C. of n independent vectors if $n \leqq 4$. We note that if $V_1, \cdots, V_n$ are independent and we take $U = 0$, then Construction 2 solves the problem of finding a shortest non-zero I. L. C. of the V's. However, Construction 2 is not a generalization of the Euclidean algorithm. Moreover, and this is vital, Construction 2 is much more laborious to apply than the algorithm which we will present.

We shall have more applications for our algorithm, and will not increase the technical difficulties in the least, if we use a generalized definition of length. Let A be a positive definite symmetric matrix.[1] We define the length (length relative to A) of U to be $(U^TAU)^{1/2}$. This length will be denoted by $L(U)$ or $L_A(U)$. When A is the unit matrix, this reduces to the customary definition of length. Note that $L(U) \geqq 0$, equality occurring if and only if $U = 0$.

Consider the problem of finding a shortest vector of the form $U - nV$. For this purpose, we seek to minimize $(U - nV)^TA(U - nV)$, which equals $U^TAU - 2nU^TAV + n^2V^TAV$. By plotting the parabola $y = U^TAU - 2xU^TAV + x^2V^TAV$, we see that we must take n a nearest integer to $(U^TAV)/(V^TAV)$. This amounts to choosing an n which minimizes $| U^TAV - nV^TAV |$. In other words, the n's which minimize the length of $U - nV$ are just the quotients which one would use to get a least numerical remainder upon dividing U^TAV by V^TAV.

The operation of replacing U by a shortest vector of the form $U - nV$ is clearly a generalization of the basic step in the one dimensional Euclidean algorithm. It will be our fundamental step, so we give it a name in Definition 3 (below). However, it alone is not quite adequate, and the other step which we use is given a name in Definition 4 (below).

[1] See M. Bôcher, *Introduction to Higher Algebra*, p. 150. By positive definite we shall mean what Bôcher would call positive definite and non-singular. So by the corollary on p. 153 of Bôcher, $U^TAU = 0$ if and only if $U = 0$.

DEFINITION 3. If some vector of the form $U - nV$ is shorter than U, and if we replace U by a shortest vector of the form $U - nV$, we say that we minimize U by integral use of V.

DEFINITION 4. If some vector of the form $U \pm V_1 \pm V_2 \pm \cdots \pm V_n$ is shorter than U, and if we replace U by a shortest vector of the form $U \pm V_1 \pm V_2 \pm \cdots \pm V_n$, we say that we minimize U by unit use of $V_1, V_2, \cdots, V_n$.

We are now ready to describe the algorithm. The algorithms for two, three, and four dimensions consist of repeated applications of Operations II, III, and IV, which we define below.

DEFINITION 5. Operation II is an operation on two vectors, and consists of the following. First, arrange the two vectors in order of increasing length. Let them then be U and V, so that $L(U) \leqq L(V)$. Then, minimize V by integral use of U.

Note that one cannot perform Operation II on just any pair of vectors. Notice further that when Operation II can be performed on two vectors, the result is a new pair of vectors, one of which is just the shorter of the two original vectors. Note also that each new vector is an I. L. C. of the old vectors and that each old vector is an I. L. C. of the new vectors.

Similar remarks will apply to Operations III and IV.

DEFINITION 6. Operation III is an operation on three vectors and consists of the following. First, arrange the three vectors in order of increasing length. Let them then be U, V, W, so that $L(U) \leqq L(V) \leqq L(W)$. Then perform the first of the following three operations which is possible.

OPERATION IIIA. Minimize one of V or W by integral use of U.

OPERATION IIIB. Minimize W by integral use of V.

OPERATION IIIC. Minimize W by unit use of U and V.

DEFINITION 7. Operation IV is an operation on four vectors and consists of the following. First, arrange the vectors in order of increasing length. Let them then be U, V, W, X, so that $L(U) \leqq L(V) \leqq L(W) \leqq L(X)$. Then, perform the first of the following seven operations which is possible.

OPERATION IVA. Minimize one of V, W, or X by integral use of U.

OPERATION IVB. Minimize one of W or X by integral use of V.

OPERATION IVC. Minimize X by integral use of W.

OPERATION IVD. Minimize one of W or X by unit use of U and V.

OPERATION IVE. Minimize X by unit use of U and W.

OPERATION IVF. Minimize X by unit use of V and W.

OPERATION IVG. Minimize X by unit use of U, V, and W.

DEFINITION 8. We shall say that $V_1, V_2, \cdots, V_n$, in the order named, is a minimal G. C. F. of $U_1, U_2, \cdots, U_m$ if the V's constitute a G. C. F. of the U's and if moreover:

1. Of all non-zero I. L. C.'s of the U's, V_1 is a shortest.
2. Of all I. L. C.'s of the U's which are independent of V_1, V_2 is a shortest.
3. Of all I. L. C.'s of the U's which are independent of V_1 and V_2, V_3 is a shortest.

And so on up to n.

In five or more dimensions, a set of vectors can have a G. C. F. without having a minimal G. C. F. To see this consider the vectors

$$U = (1, 0, 0, 0, 0),$$
$$V = (0, 1, 0, 0, 0),$$
$$W = (0, 0, 1, 0, 0),$$
$$X = (0, 0, 0, 1, 0),$$
$$Y = (0, 0, 0, 0, 1),$$
$$Z = (\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2})$$

under the usual definition of length. The last and any four of the first five will constitute a G. C. F., but there is no minimal G. C. F. On the other hand, any commensurable set of vectors in four or less dimensions will have a minimal G. C. F. We will show that, given a G. C. F. in four or less dimensions, one can derive a minimal G. C. F. from it by successive applications of Operations II, III, or IV.

We first need some theorems concerning quadratic forms, so we let $f(x, y)$, $f(x, y, z)$, $\cdots$ denote quadratic forms in two, three, $\cdots$ variables. Also, it will be necessary to break the argument up into cases according as the various variables are positive or negative. In order to do this expeditiously, we will assume the variables always non-negative and multiply x by e_1, y by e_2, $\cdots$, where each of e_1, e_2, $\cdots$ can be either 1 or -1.

THEOREM 6. *Let*

$$0 < f(1, 0) \leqq f(0, 1), \tag{1}$$

$$f(0, 1) \leqq f(e_1, e_2). \tag{2}$$

Let x and y not both be zero, and let α be the least non-zero one of x and y. Then

$$\alpha^2 f(1, 0) \leqq f(e_1 x, e_2 y), \tag{3}$$

and if $y \neq 0$, then

$$\alpha^2 f(0, 1) \leqq f(e_1 x, e_2 y). \tag{4}$$

Moreover if neither of x and y is zero and if x and y are unequal, then the "$\leqq$" in (4) can be replaced by "$<$".

Proof. Clearly, if $y = 0$, (3) is true. If $y \neq 0$, then (3) follows from (4) and (1). So we undertake to prove (4). If $x = 0$, (4) is clearly true. So let $x \neq 0$ and $y \neq 0$.

Case 1. $x \geqq y$. If $g(x, y)$ is a quadratic form, then

$$g(x, y) = g(1, 1)y^2 + g(1, 0)(x - y)^2 + g(1, 0)y(x - y) + \{g(1, 1) - g(0, 1)\}y(x - y).$$

To see this, put $g(x, y) \equiv u_1x^2 + u_2xy + u_3y^2$ and multiply out the right side of the equation. Putting $g(x, y) = f(e_1x, e_2y)$, we get

$$f(e_1x, e_2y) = f(e_1, e_2)y^2 + f(e_1, 0)(x - y)^2 + f(e_1, 0)y(x - y) + \{f(e_1, e_2) - f(0, e_2)\}y(x - y).$$

Now by (2)

$$f(e_1, e_2)y^2 \geqq f(0, 1)y^2 = f(0, 1)\alpha^2.$$

Also

$$f(e_1, 0) = f(1, 0) > 0.$$

Also by (2),

$$f(e_1, e_2) - f(0, e_2) = f(e_1, e_2) - f(0, 1) \geqq 0.$$

So

$$f(e_1x, e_2y) \geqq f(0, 1)\alpha^2,$$

and if $x \neq y$,

$$f(e_1x, e_2y) > f(0, 1)\alpha^2.$$

Case 2. $x \leqq y$. In this case, put $g(y, x) = f(e_1x, e_2y)$. So

$$f(e_1x, e_2y) = f(e_1, e_2)x^2 + f(0, e_2)(y - x)^2 + f(0, e_2)x(y - x) + \{f(e_1, e_2) - f(e_1, 0)\}x(y - x).$$

However, by (2),

$$f(e_1, e_2) \geqq f(0, 1)$$

and by (1),

$$f(0, 1) \geqq f(1, 0) = f(e_1, 0).$$

So

$$f(e_1, e_2) - f(e_1, 0) \geqq 0.$$

So

$$f(e_1x, e_2y) \geqq f(0, 1)\alpha^2,$$

and if $x \neq y$,

$$f(e_1x, e_2y) > f(0, 1)\alpha^2.$$

THEOREM 7. *Let*

(1) $$0 < f(1, 0, 0) \leqq f(0, 1, 0) \leqq f(0, 0, 1),$$

$$f(0, 1, 0) \leqq f(e_1 , e_2 , 0), \tag{2}$$

$$f(0, 0, 1) \leqq f(e_1 , 0, e_3), \tag{3}$$

$$f(0, 0, 1) \leqq f(0, e_2 , e_3), \tag{4}$$

$$f(0, 0, 1) \leqq f(e_1 , e_2 , e_3). \tag{5}$$

Let x, y, and z not all be zero, and let α be the least non-zero one of x, y, and z. Then

$$\alpha^2 f(1, 0, 0) \leqq f(e_1x, e_2y, e_3z), \tag{6}$$

and if $y \neq 0$,

$$\alpha^2 f(0, 1, 0) \leqq f(e_1x, e_2y, e_3z), \tag{7}$$

and if $z \neq 0$,

$$\alpha^2 f(0, 0, 1) \leqq f(e_1x, e_2y, e_3z). \tag{8}$$

Moreover, if none of x, y, or z is zero, and x, y, and z are not all equal, then the "$\leqq$" in (8) can be replaced by a "$<$".

Proof. If y and z are both zero, then (6) holds. If $y \neq 0$, then (6) follows from (7) and (1). If $z \neq 0$, then (6) follows from (8) and (1). So we turn to (7) and (8). If $z = 0$, we may consider $f(x, y, z)$ as a quadratic form in x and y only, and so (7) follows by Theorem 6. If $z \neq 0$, then (7) follows from (8) and (1). So we have only to prove (8). If either of x or y is zero, we may consider $f(x, y, z)$ as a quadratic form in the other two variables and apply Theorem 6. So let $x \neq 0$, $y \neq 0$, $z \neq 0$.

Case 1. $x = y \geqq z$. Let $g(x, z) = f(e_1x, e_3x, e_3z) = f(e_1x, e_2y, e_3z)$. Then

$$f(e_1x, e_2y, e_3z) = f(e_1 , e_2 , e_3)z^2 + f(e_1 , e_2 , 0)(x - z)^2 + f(e_1 , e_2 , 0)z(x - z) + \{f(e_1 , e_2 , e_3) - f(0, 0, e_3)\}z(x - z).$$

So by (5) and (2),

$$f(e_1x, e_2y, e_3z) \geqq f(0, 0, 1)\alpha^2.$$

Case 2 and *Case* 3 are $x = z \geqq y$ and $y = z \geqq x$, and are handled similarly.

Case 4. $x > y \geqq z$. If $g(x, y, z)$ is a quadratic form, then

$$g(x, y, z) = g(y, y, z) + \{g(1, 1, 0) - g(0, 1, 0)\}y(x - y) + \{g(1, 0, 1) - g(0, 0, 1)\}z(x - y) + g(1, 0, 0)(x - y)(y - z) + g(1, 0, 0)(x - y)^2.$$

If we put $g(x, y, z) = f(e_1x, e_2y, e_3z)$, we get

$$f(e_1x, e_2y, e_3z) = f(e_1y, e_2y, e_3z) + \{f(e_1 , e_2 , 0) - f(0, e_2 , 0)\}y(x - y) + \{f(e_1 , 0, e_3) - f(0, 0, e_3)\}z(x - y) + f(e_1 , 0, 0)(x - y)(y - z) + f(e_1 , 0, 0)(x - y)^2.$$

By Case 1, $f(e_1y, e_2y, e_3z) \geqq f(0, 0, 1)\alpha^2$.

Cases 5–9, namely, $y > x \geqq z$, $x > z \geqq y$, $z > x \geqq y$, $y > z \geqq x$, and $z > y \geqq x$ are handled similarly.

THEOREM 8. *Let*

$$0 < f(1, 0, 0, 0) \leqq f(0, 1, 0, 0) \leqq f(0, 0, 1, 0) \leqq f(0, 0, 0, 1), \tag{1}$$

$$f(0, 1, 0, 0) \leqq f(e_1 , e_2 , 0, 0), \tag{2}$$

$$f(0, 0, 1, 0) \leqq f(e_1 , 0, e_3 , 0), \tag{3}$$

$$f(0, 0, 1, 0) \leqq f(0, e_2 , e_3 , 0), \tag{4}$$

$$f(0, 0, 1, 0) \leqq f(e_1 , e_2 , e_3 , 0), \tag{5}$$

$$f(0, 0, 0, 1) \leqq f(e_1 , 0, 0, e_4), \tag{6}$$

$$f(0, 0, 0, 1) \leqq f(0, e_2 , 0, e_4), \tag{7}$$

$$f(0, 0, 0, 1) \leqq f(0, 0, e_3 , e_4), \tag{8}$$

$$f(0, 0, 0, 1) \leqq f(e_1 , e_2 , 0, e_4), \tag{9}$$

$$f(0, 0, 0, 1) \leqq f(e_1 , 0, e_3 , e_4), \tag{10}$$

$$f(0, 0, 0, 1) \leqq f(0, e_2 , e_3 , e_4), \tag{11}$$

$$f(0, 0, 0, 1) \leqq f(e_1 , e_2 , e_3 , e_4). \tag{12}$$

Let x, y, z, and w not all be zero, and let α be the least non-zero one of $x, y, z,$ and w. If some of x, y, z, w are not integers, put $\beta = 3\alpha^2/4$. If all of x, y, z, w are integers, put $\beta = \max(1, 3\alpha^2/4)$. Then

$$\beta f(1, 0, 0, 0) \leqq f(e_1x, e_2y, e_3z, e_4w), \tag{13}$$

and if $y \neq 0$,

$$\beta f(0, 1, 0, 0) \leqq f(e_1x, e_2y, e_3z, e_4w), \tag{14}$$

and if $z \neq 0$,

$$\beta f(0, 0, 1, 0) \leqq f(e_1x, e_2y, e_3z, e_4w), \tag{15}$$

and if $w \neq 0$,

$$\beta f(0, 0, 0, 1) \leqq f(e_1x, e_2y, e_3z, e_4w). \tag{16}$$

Moreover, if none of x, y, z, or w is zero, and if there are at least three different values among the x, y, z, and w, then the "$\leqq$" in (16) can be replaced by a "$<$".

Proof. As in the proof of Theorem 7, we may confine our attention to the case where $x \neq 0$, $y \neq 0$, $z \neq 0$, $w \neq 0$, and we are seeking to prove (16). We first prove a lemma.

LEMMA. *If conditions* (1)–(12) *are satisfied, and none of x, y, z, or w is zero, and if the two largest of x, y, z, and w are equal and if α is the value of the smallest of x, y, z, and w, then*

$$\alpha^2 f(0, 0, 0, 1) \leqq f(e_1x, e_2y, e_3z, e_4w).$$

Moreover, if the x, y, z, and w are not all equal, one can replace the "$\leqq$" by a "$<$".

Case 1. $x = y = z \geqq w$. Put

$$g(x, w) = f(e_1x, e_2x, e_3x, e_4w) = f(e_1x, e_2y, e_3z, e_4w).$$

Then

$$\begin{aligned} &f(e_1x, e_2y, e_3z, e_4w) \\ &\quad = f(e_1, e_2, e_3, e_4)w^2 + f(e_1, e_2, e_3, 0)(x - w)^2 + f(e_1, e_2, e_3, 0)w(x - w) \\ &\quad + \{f(e_1, e_2, e_3, e_4) - f(0, 0, 0, e_4)\}w(x - w) \geqq f(0, 0, 0, 1)\alpha^2. \end{aligned}$$

Cases 2, 3, and 4, namely, $x = y = w \geqq z$, $x = z = w \geqq y$, and $y = z = w \geqq x$ proceed similarly.

Case 5. $x = y > z \geqq w$. Put $g(x, z, w) = f(e_1x, e_2x, e_3z, e_4w) = f(e_1x, e_2y, e_3z, e_4w)$. Then

$$\begin{aligned} f(e_1x, e_2y, e_3z, e_4w) &= f(e_1z, e_2z, e_3z, e_4w) \\ &\quad + \{f(e_1, e_2, e_3, 0) - f(0, 0, e_3, 0)\}z(x - z) \\ &\quad + \{f(e_1, e_2, 0, e_4) - f(0, 0, 0, e_4)\}w(x - z) \\ &\quad + f(e_1, e_2, 0, 0)(x - z)(z - w) + f(e_1, e_2, 0, 0)(x - z)^2 \\ &\quad \geqq f(0, 0, 0, 1)\alpha^2 \text{ by Case 1.} \end{aligned}$$

Cases 6–16, namely, $x = z > y \geqq w$, $y = z > x \geqq w$, $x = y > w \geqq z$, $x = w > y \geqq z$, $y = w > x \geqq z$, $x = w > z \geqq y$, $x = z > w \geqq y$, $w = z > x \geqq y$, $y = z > w \geqq x$, $y = w > z \geqq x$, $z = w > y \geqq x$ proceed similarly.

This completes the proof of the lemma. We return to the proof of the main theorem, noting that the lemma disposes of all cases except where the two largest of x, y, z, and w are unequal.

Case 1. $x > y \geqq z \geqq w$. If $g(x, y, z, w)$ is a quadratic form, then

$$\begin{aligned} g(x, y, z, w) &= g(y, y, z, w) + \{g(1, 1, 0, 0) - g(0, 1, 0, 0)\}y(x - y) \\ &\quad + \{g(1, 0, 1, 0) - g(0, 0, 1, 0)\}z(x - y) \\ &\quad + \{g(1, 0, 0, 1) - g(0, 0, 0, 1)\}w(x - y) \\ &\quad + g(1, 0, 0, 0)(x - y)(y - z) + g(1, 0, 0, 0)(x - y - w)(x - y). \end{aligned}$$

Putting $g(x, y, z, w) = f(e_1x, e_2y, e_3z, e_4w)$, we get

$$\begin{aligned} f(e_1x, e_2y, e_3z, e_4w) &= f(e_1y, e_2y, e_3z, e_4w) \\ &\quad + \{f(e_1, e_2, 0, 0) - f(0, e_2, 0, 0)\}y(x - y) \\ &\quad + \{f(e_1, 0, e_3, 0) - f(0, 0, e_3, 0)\}z(x - y) \end{aligned}$$

$$+ \{f(e_1, 0, 0, e_4) - f(0, 0, 0, e_4)\}w(x - y)$$
$$+ f(e_1, 0, 0, 0)(x - y)(y - z)$$
$$+ f(e_1, 0, 0, 0)(x - y - w)(x - y).$$

By the lemma this $\geqq$

$$f(0, 0, 0, 1)w^2 + f(1, 0, 0, 0)(x - y - w)(x - y).$$

Also, if there are at least three different values among x, y, z, and w, then y, z, and w are not all equal, and we may replace the "$\geqq$" by a "$>$". If $x - y \geqq w$, our theorem is proved. So let $x - y < w$. Now $f(1, 0, 0, 0) \leqq f(0, 0, 0, 1)$. Also $(w - u)u$ takes its maximum at $u = \frac{1}{2}w$, and so we get $(w - u)u \leqq \frac{1}{4}w^2$. So

$$(w - (x - y))(x - y) \leqq \tfrac{1}{4}w^2.$$

So

$$f(0, 0, 0, 1)w^2 + f(1, 0, 0, 0)(x - y - w)(x - y)$$
$$\geqq f(0, 0, 0, 1)w^2 - f(0, 0, 0, 1)\tfrac{1}{4}w^2$$
$$\geqq \tfrac{3}{4}\alpha^2 f(0, 0, 0, 1).$$

If some of x, y, z, w are not integers, then $\beta = 3\alpha^2/4$, and we have proved

$$f(e_1x, e_2y, e_3z, e_4w) \geqq \beta f(0, 0, 0, 1).$$

Suppose all of x, y, z, w are integers. As before, we have

$$(w - (x - y))(x - y) \leqq \tfrac{1}{4}w^2.$$

So

$$(w - (x - y))(x - y) < w^2.$$

Since we are now dealing with integers,

$$(w - (x - y))(x - y) \leqq w^2 - 1.$$

So

$$(w - (x - y))(x - y) \leqq \min\,(w^2 - 1, \tfrac{1}{4}w^2).$$

So

$$f(0, 0, 0, 1)w^2 + f(1, 0, 0, 0)(x - y - w)(x - y)$$
$$\geqq f(0, 0, 0, 1)w^2 - f(0, 0, 0, 1)\,\min\,(w^2 - 1, \tfrac{1}{4}w^2)$$
$$\geqq f(0, 0, 0, 1)\,\max\,(1, \tfrac{3}{4}w^2).$$

The remaining cases of the theorem proceed similarly.

Note one point about the preceding three theorems. As they are stated (and

proved), $e_1, e_2, \cdots$ could denote a particular choice of signs for $x, y, \cdots$. If so, then the theorem says that if the hypothesis is true for a particular choice of signs, then the conclusion is true for the same choice of signs. From this it clearly follows that if the hypothesis is true for all possible choices (as will be the case in the application), then the conclusion is likewise true for all choices.

Now suppose U_1, V_1 is a G. C. F. of $R_1, \cdots, R_m$, and suppose one can apply Operation II to U_1, V_1, getting U_2, V_2. Then U_2, V_2 is also a G. C. F. of $R_1, \cdots, R_m$. Also it is clear from the nature of Operation II, that $L(U_2) + L(V_2) < L(U_1) + L(V_1)$. Suppose now that Operation II can be performed on U_2, V_2 also, giving U_3, V_3. Then U_3, V_3 is a G. C. F. of $R_1, \cdots, R_m$ and $L(U_3) + L(V_3) < L(U_2) + L(V_2)$. Can we continue to perform Operation II indefinitely? No, for the following reasons. Since $R_1, \cdots, R_m$ have a G. C. F., they are commensurable (this can easily be proved for the generalized length by noting that, since $A = P^TP$, then $[L_A(U)]^2 = U^TAU = U^TP^TPU = (PU)^T(PU) =$ the square of the ordinary length of PU. In other words, the usual definition of length is restored by transforming by use of P). If there are an infinite succession of G. C. F.'s of the R's, namely U_n, V_n for $n = 1, 2, \cdots$, having the property that $L(U_{n+1}) + L(V_{n+1}) < L(U_n) + L(V_n)$, then there would have to be an infinite number of different I. L. C.'s of the R's in a bounded region. These would have to have a limit point, contrary to the fact that the R's are commensurable. Hence if we start with a G. C. F. of the R's and keep performing Operation II successively, we will come at last to a G. C. F. of the R's on which Operation II cannot be performed. When the vectors of this are arranged in order of increasing length, they will constitute a minimal G. C. F. of the R's, as we prove in the theorem below.

THEOREM 9. *Suppose U, V is a* G. C. F. *of $R_1, \cdots, R_m$, and $L(U) \leqq L(V)$, and Operation* II *cannot be performed on U, V. Then U, V is a minimal* G. C. F. *of $R_1, \cdots, R_m$.*

Proof. Since U, V is a G. C. F., $U \neq 0$. Hence $0 < L(U)$. Also, since Operation II cannot be performed, $L(V) \leqq L(V - nU)$. In particular, $L(V) \leqq L(V \pm U)$. As $L(W) = L(-W)$, $L(V) \leqq L(\pm V \pm U)$. So

$$0 < [L(U)]^2 \leqq [L(V)]^2, \tag{1}$$

$$[L(V)]^2 \leqq [L(\pm V \pm U)]^2. \tag{2}$$

Now put

$$f(x, y) = x^2U^TAU + 2xyU^TAV + y^2V^TAV$$
$$= (xU + yV)^TA(xU + yV) = [L(xU + yV)]^2.$$

Then (1) and (2) become

$$0 < f(1, 0) \leqq f(0, 1),$$
$$f(0, 1) \leqq f(e_1, e_2),$$

where the second condition holds for all possible determinations of the ϵ's. So we have the hypothesis of Theorem 6 satisfied. Now any I. L. C. of the R's is an I. L. C. of U, V. So let $aU + bV$ be a shortest non-zero I. L. C. of the R's. Then $[L(aU + bV)]^2 = f(a, b)$. By Theorem 6, $f(a, b) \geqq f(1, 0)$. As $f(1, 0) = [L(U)]^2$, we have $L(aU + bV) \geqq L(U)$. But $aU + bV$ is a shortest non-zero I. L. C. of the R's, and U is a non-zero I. L. C. of the R's. So U is a shortest non-zero I. L. C. of the R's. Now let $cU + dV$ be a shortest I. L. C. of the R's which is independent of U. Then $d \neq 0$. So by Theorem 6, $[L(cU + dV)]^2 = f(c, d) \geqq f(0, 1) = [L(V)]^2$. So V is a shortest I. L. C. of the R's which is independent of U. So U, V is a minimal G. C. F. of the R's.

Suppose U, V, W is a G. C. F. of a set of vectors $R_1, \cdots, R_m$ and one successively performs Operation III on U, V, W. By an argument similar to that used above, there must come a time when Operation III can no longer be performed. If we then arrange the three resultant vectors in order of increasing length, they will constitute a minimal G. C. F. of the R's, as the following theorem shows.

THEOREM 10. *Suppose U, V, W is a* G. C. F. *of $R_1, \cdots, R_m$, and $L(U) \leqq L(V) \leqq L(W)$, and Operation* III *cannot be performed on U, V, W. Then U, V, W is a minimal* G. C. F. *of $R_1, \cdots, R_m$.*

The proof is similar to the proof of Theorem 9. If we put $f(x, y, z) = [L(xU + yV + zW)]^2$, then $0 < L(U) \leqq L(V) \leqq L(W)$, and this together with the inequalities derivable from the fact that Operation III cannot be performed insure that the hypothesis of Theorem 7 is satisfied. The conclusion of Theorem 7 can be used as the conclusion of Theorem 6 was used in the proof of Theorem 9, to show that U, V, W is a minimal G. C. F.

If U, V, W, X is a G. C. F. of $R_1, \cdots, R_m$, then one can find a minimal G. C. F. of the R's by use of Operation IV, as the following theorem shows.

THEOREM 11. *Suppose U, V, W, X is a* G. C. F. *of $R_1, \cdots, R_m$, and $L(U) \leqq L(V) \leqq L(W) \leqq L(X)$, and Operation* IV *cannot be performed on U, V, W, X. Then U, V, W, X is a minimal* G. C. F. *of $R_1, \cdots, R_m$.*

The proof is similar to the proof of Theorem 9.

We have now shown that if one has a G. C. F. in four or less dimensions, then one can find a minimal G. C. F. by use of Operations II, III, and IV. The question of how to find a G. C. F. in the first place is relatively unimportant, for in all the applications that I have found for the algorithm so far, either one has a G. C. F. given to start with, or else there is no G. C. F. However, it could be shown that if one should start with three dependent commensurable vectors, not all zero, and apply Operation III, one would eventually find a G. C. F. Also, it could be shown that if one should start with four dependent commensurable vectors, not all zero, and apply Operation IV, one would eventually find a G. C. F. Whether one could find a G. C. F. for five dependent commensurable vectors by use of an "Operation V" obtained by strict generalization of Operations II, III, and IV, I do not know. Certainly nothing would be

lost by performing such an "Operation V" on the five vectors as long as it could be performed, and saving the more complicated earlier methods (see Theorem 5) as a last resort.

We now consider a number of applications.

PROBLEM I. Let us ask for the minimum (different from zero, of course) of the positive definite quadratic form

$$Q(a, b, c, d) = 7a^2 + 14b^2 + 326c^2 + 81d^2 + 18ab + 90ac + 45ad + 134bc + 66bd + 324cd.$$

If we put

$$A = \begin{Vmatrix} 14 & 18 & 90 & 45 \\ 18 & 28 & 134 & 66 \\ 90 & 134 & 652 & 324 \\ 45 & 66 & 324 & 162 \end{Vmatrix}$$

and $V = (a, b, c, d)$, then $V^TAV = 2Q(a, b, c, d)$. So we wish to minimize V^TAV. If we put

$$V_1 = (1, 0, 0, 0),$$
$$V_2 = (0, 1, 0, 0),$$
$$V_3 = (0, 0, 1, 0),$$
$$V_4 = (0, 0, 0, 1),$$

and base our definition of length on A, then we wish to find a shortest non-zero I. L. C. of V_1, V_2, V_3, V_4. For this, it suffices to find a minimal G. C. F. of V_1, V_2, V_3, V_4. As V_1, V_2, V_3, V_4 are independent, they constitute a G. C. F. of themselves. Hence, if we apply Operation IV to V_1, V_2, V_3, V_4, we will arrive at a minimal G. C. F.

In order to illustrate the ease with which the applications of Operation IV can be performed, we have prepared a table summarizing the computations. It is Table I at the end of the paper. The values of $V_i^TAV_j$ are useful in performing Operation IV, so in Table I we have listed the V's along the top and left side, and have put the value of $V_i^TAV_j$ under V_i and to the right of V_j. For $1 \leqq i \leqq 4$ and $1 \leqq j \leqq 4$, the value of $V_i^TAV_j$ is just the component of A in the i-th row and j-th column. Note that V_1 is the shortest of the four initial vectors. So for Operation IVA, we seek to minimize one of V_2, V_3, V_4 by integral use of V_1. Specifically, we pick on V_2. The n which minimizes $L(V_2 - nV_1)$ is just the n which minimizes $| V_2^TAV_1 - nV_1^TAV_1 |$, as we showed earlier. By inspection, this n is unity. So we define $V_5 = V_2 - V_1$, and replace V_2 by V_5. The values of $V_i^TAV_5$ can be readily computed by noting that

$$V_i^T A(V_j - nV_k) = V_i^T AV_j - nV_i^T AV_k .$$

So the numbers under V_1, V_3, and V_4 and to the right of V_5 are obtained by subtracting the numbers to the right of V_1 from the numbers to the right of V_2. The values of $V_5^T AV_i$ can be filled in immediately from the relation $V_i^T AV_j = V_j^T AV_i$. Finally $V_5^T AV_5$ can be computed from the relation

$$(V_i - nV_j)^T A(V_i - nV_j) = V_i^T AV_i - 2nV_i^T AV_j + n^2 V_j^T AV_j .$$

Note that we do not record $V_2^T AV_5$ or $V_5^T AV_2$ as we are discarding V_2 in favor of V_5. Now V_5 is shorter than each of V_1, V_3, and V_4, so that our next three uses of Operation IV are uses of Operation IVA, in which we successively minimize V_1, V_3, and V_4 by integral use of V_5. Then for V_9, we minimize V_7 by integral use of V_6, which is an application of Operation IVB. For V_{10}, we return to Operation IVA again. Our first use of Operation IVC comes in finding V_{13}, after which we again return to Operation IVA. We observe that V_{15} gives Q the value unity, which is clearly a minimum, so we stop, although we do not yet have a minimal G. C. F. So $a = 1, b = 2, c = -1, d = 1$ makes $Q = 1$, a minimum.

This minimum is not unique, because there are other sets of values of a, b, c, d which also give Q the value unity. After we have considered the problem of finding all minimal G. C. F.'s we will find all sets of a, b, c, d which make $Q = 1$.

We now prepare the way for Problem II by a short discussion.

A well-known theorem[2] states:

Given n linear forms in n variables with arbitrary real coefficients and a determinant $\Delta \neq 0$, one can choose integer values of the variables, not all zero, such that, simultaneously, each form $\leqq |\Delta|^{1/n}$ in absolute value.

Naturally this raises the question of how to find the integers whose existence is stated. In the book *Minkowski Geometry of Numbers* by Harris Hancock, pp. 378–444 are devoted to a description and discussion of an algorithm for finding the integers in the case $n = 3$. No information is given for the case $n = 4$. Let us consider the four forms

$$L_1(x, y, z, w) = 831x + 1242y + 913z + 502w,$$

$$L_2(x, y, z, w) = 799x + 1401y + 288z + 311w,$$

$$L_3(x, y, z, w) = 992x + 411y + 776z + 168w,$$

$$L_4(x, y, z, w) = 321x + 889y + 1706z + 842w.$$

For these forms, $\Delta = 255$. So $\Delta^{1/4} < 4$. So there must be integer values of x, y, z, and w which make $|L_i| \leqq 3$ simultaneously.

PROBLEM II. Find integer values of x, y, z, w, not all zero, such that $|L_i| \leqq 3$ simultaneously.

[2] H. Minkowski, *Geometrie der Zahlen*, p. 104.

Putting

$$V_1 = (831, 799, 992, 321),$$
$$V_2 = (1242, 1401, 411, 889),$$
$$V_3 = (913, 288, 776, 1706),$$
$$V_4 = (502, 311, 168, 842),$$

we see that Problem II is equivalent to the problem of finding integers x, y, z, w such that the components of $xV_1 + yV_2 + zV_3 + wV_4$ are all $\leqq 3$ in absolute value. Clearly this vector $xV_1 + yV_2 + zV_3 + wV_4$ has length (in the usual sense) $\leqq 6$. So if we can find all I. L. C.'s of V_1, V_2, V_3, and V_4 which have length $\leqq 6$, one of them must be the vector $xV_1 + yV_2 + zV_3 + wV_4$. Presumably this problem could be solved by use of Construction 1 (with $U = 0$). However, to apply Construction 1 to V_1, V_2, V_3, V_4 directly would be impracticable because there would be an enormous number of cases to try, each involving a most unpleasant amount of computation. Let us instead first find a minimal G. C. F. of V_1, V_2, V_3, V_4. This can be done rather quickly and we get the following for a minimal G. C. F.

$$U_1 = (0, 1, 2, 1),$$
$$U_2 = (-1, 3, -2, 0),$$
$$U_3 = (-3, -1, 1, -2),$$
$$U_4 = (-4, -4, -1, 5).$$

Already we have three I. L. C.'s of the V's with all their components $\leqq 3$ in absolute value, namely U_1, U_2, U_3. A slight bit of trial and error discloses two more, namely

$$U_3 + U_1 = (-3, 0, 3, -1),$$
$$U_3 - U_1 = (-3, -2, -1, -3).$$

These are probably all, but one could be sure by applying Construction 1 to find all I. L. C.'s of the U's with length $\leqq 6$, which could be done rather quickly. Of course we still have to exhibit an actual set of values of x, y, z, w. So let us find the x, y, z, w such that $xV_1 + yV_2 + zV_3 + wV_4 = U_1$. This merely amounts to solving the four equations $L_1 = 0, L_2 = 1, L_3 = 2, L_4 = 1$ simultaneously. We record the answer, together with the four other answers got by putting $xV_1 + yV_2 + zV_3 + wV_4$ equal respectively to U_2, U_3, $U_3 + U_1$, and $U_3 - U_1$ in Table II at the end of the paper.

We prepare the way for Problem III with certain explanatory remarks.

A well-known theorem[3] states:

[3] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, p. 169.

If $x_1, \cdots, x_n$ are real numbers, not all rational, then the system of simultaneous inequalities on the a's,

$$\text{(A)} \qquad |ax_i - a_i| < a^{-1/n} \qquad (i = 1, 2, \cdots, n)$$

has an infinity of solutions in integers.

However, there is no well-known technique for finding such solutions.[4] We shall propose a technique, but we are unable to prove that it always works. However, it has never yet failed when applied to a particular numerical problem, and we shall cite one severe test which it survived. To exhibit the capabilities of the technique, we shall solve a slightly more difficult problem than merely finding some random solutions of the inequalities (A).

PROBLEM III. Prove, by exhibiting the necessary a's, b's, and c's, that for each z between 1 and 1,000,000 inclusive there are positive integers a, b, and c such that $a \leqq z$ and

$$|a\pi - b| < z^{-1/2},$$

$$|a\gamma - c| < z^{-1/2}.$$

Note that since $a \leqq z$, the a, b, and c satisfy (A) with $x_1 = \pi$ and $x_2 = \gamma$.

In order to solve this problem it is necessary to find, for various values of ϵ, values of a, b, and c such that

$$|a\pi - b| < \epsilon,$$

$$|a\gamma - c| < \epsilon.$$

If we put $V_1 = (\pi, \gamma)$, $V_2 = (-1, 0)$, and $V_3 = (0, -1)$, then the coordinates of $aV_1 + bV_2 + cV_3$ are $a\pi - b$ and $a\gamma - c$. So we wish to find short I. L. C.'s of V_1, V_2, V_3. To this end we apply Operation III repeatedly. By making a judicious selection from the shorter and shorter I. L. C.'s of V_1, V_2, V_3 which we obtain by use of Operation III, we were able to make up Table III at end of paper, which solves Problem III by exhibiting sets of values of a, b, c, each of which satisfies the desired inequalities for a given range of z, and such that the z-ranges overlap and cover the interval from 1 to 1,000,000.

As a test of the technique, I undertook to find solutions of (A) when $x_1 = \log 3/\log 2$, $x_2 = \log 5/\log 2$, $x_3 = \log 7/\log 2$, $x_4 = \log 17/\log 2$. Here one would put

$$V_1 = (x_1, x_2, x_3, x_4),$$

$$V_2 = (-1, 0, 0, 0),$$

$$V_3 = (0, -1, 0, 0),$$

$$V_4 = (0, 0, -1, 0),$$

$$V_5 = (0, 0, 0, -1)$$

[4] When $n = 1$, the solutions can be found by continued fractions. However Jacobi's generalized continued fraction algorithm fails to furnish solutions for $n > 1$ except for special cases. See O. Perron, *Grundlagen für eine Theorie des Jacobischen Kettenbruchalgorithmus*, Math. Ann., vol. 64(1907), pp. 1–76, especially pp. 20–23.

and search for short I. L. C.'s of the V's. To carry out the search, one would apply a generalization of Operations II, III, and IV. This generalized operation was very successful in finding solutions of (A), and continued to be successful as far as the computation was carried. In order to make the test a severe one, the computation was carried quite a way, so far in fact, that it was necessary to use fifteen decimal place approximations for the logarithms toward the end of the computation. A sample solution of (A) (this was obtained near the end of the computation) is $a = 60{,}243{,}195$, $a_1 = 95{,}483{,}205$, $a_2 = 139{,}880{,}367$, $a_3 = 169{,}124{,}030$, $a_4 = 246{,}241{,}821$. Note that this is not merely a solution of (A) but satisfies the stronger inequalities

$$| ax_i - a_i | < \tfrac{1}{3}a^{-1/4} \qquad (i = 1, 2, 3, 4).$$

We postpone further applications until after more development of the theory.

THEOREM 12. *If U_1, V_1 is a minimal* G. C. F. *of U, V, then U_1 is at least as short as the shorter of U and V, and V_1 is at least as short as the longer of U and V.*

Proof. By definition, U_1 is a shortest non-zero I. L. C. of U and V. Now U and V are independent, since otherwise their G. C. F. would consist of a single vector. So both U and V are non-zero I. L. C.'s of U and V. Hence U_1 must be at least as short as the shorter of U and V. Since U and V are independent, they are not both dependent on U_1. Hence at least one of them is an I. L. C. of U and V which is independent of U_1, and so must be at least as long as V_1, by the definition of a minimal G. C. F. So V_1 must be at least as short as the longer one of U and V.

THEOREM 13. *If U_1, V_1, W_1 is a minimal* G. C. F. *of U, V, W, then U_1 is at least as short as the shortest of U, V, W, and V_1 is at least as short as the second shortest of U, V, W, and W_1 is at least as short as the longest of U, V, and W.*

The proof is similar to that of Theorem 12.

THEOREM 14. *If U_1, V_1, W_1, X_1 is a minimal* G. C. F. *of U, V, W, X and $L(U) \leqq L(V) \leqq L(W) \leqq L(X)$, then $L(U_1) \leqq L(U)$, $L(V_1) \leqq L(V)$, $L(W_1) \leqq L(W)$, and $L(X_1) \leqq L(X)$.*

The proof is similar.

THEOREM 15. *If U_1, V_1 and U_2, V_2 are both minimal* G. C. F.*'s of R_1, $\cdots$, R_m, then $L(U_1) = L(U_2)$ and $L(V_1) = L(V_2)$.*

Proof. Since U_1, V_1 and U_2, V_2 are both minimal G. C. F.'s of R_1, $\cdots$, R_m, it follows that U_1, V_1 is a minimal G. C. F. of U_2, V_2. So by Theorem 12, $L(U_1) \leqq L(U_2)$ and $L(V_1) \leqq L(V_2)$. Interchanging U_1, V_1 with U_2, V_2, we get $L(U_2) \leqq L(U_1)$ and $L(V_2) \leqq L(V_1)$.

THEOREM 16. *If U_1, V_1, W_1 and U_2, V_2, W_2 are both minimal* G. C. F.*'s of R_1, $\cdots$, R_m, then $L(U_1) = L(U_2)$, $L(V_1) = L(V_2)$, and $L(W_1) = L(W_2)$.*

The proof is similar.

THEOREM 17. *If* U_1, V_1, W_1, X_1 *and* U_2, V_2, W_2, X_2 *are both minimal* G. C. F.*'s of* $R_1, \cdots, R_m$, *then* $L(U_1) = L(U_2), L(V_1) = L(V_2), L(W_1) = L(W_2)$, *and* $L(X_1) = L(X_2)$.

The proof is similar.

We now make a few remarks preliminary to the statement of Problem IV.

Put $L(x, y, z, w) = 985x + 408y - 780z - 571w$. A general solution in integers of $L = k_1$ is[5]

$$\begin{aligned} x &= 5k_1 + 8k_2 + 3k_3 + 9k_4, \\ y &= 5k_1 + 11k_2 + 10k_3 + 5k_4, \\ z &= 6k_1 + 10k_2 + 20k_3 + 3k_4, \\ w &= 4k_1 + 8k_2 - 15k_3 + 15k_4, \end{aligned}$$

where k_1 is the same as in the equation $L = k_1$, and k_2, k_3, k_4 are arbitrary. If we put

$$\begin{aligned} V &= (x, y, z, w), \\ V_1 &= (5, 5, 6, 4), \\ V_2 &= (8, 11, 10, 8), \\ V_3 &= (3, 10, 20, -15). \\ V_4 &= (9, 5, 3, 15), \end{aligned}$$

then our general solution takes the form

$$V = \sum_{i=1}^{4} k_i V_i.$$

PROBLEM IV. Find a general solution in integers

$$\begin{aligned} x &= a_{11}k_1 + a_{12}k_2 + a_{13}k_3 + a_{14}k_4, \\ y &= a_{21}k_1 + a_{22}k_2 + a_{23}k_3 + a_{24}k_4, \\ z &= a_{31}k_1 + a_{32}k_2 + a_{33}k_3 + a_{34}k_4, \\ w &= a_{41}k_1 + a_{42}k_2 + a_{43}k_3 + a_{44}k_4 \end{aligned}$$

of $L = k_1$ for which $\sum(a_{ij})^2$ shall be a minimum.

An equivalent statement of the problem is clearly the following.

[5] Jacobi is usually credited with being the first to give a systematic method for finding such solutions. However in his treatment of the question (*Werke*, vol. 6, pp. 355–384) he contrasts certain of his results with results of Euler. The solution given here is obtained by the method of Rosser, *A note on the linear Diophantine equation*, Amer. Math. Monthly, vol. 48(1941), pp. 662–666.

PROBLEM IV. Find a general solution in integers

$$V = \sum_{i=1}^{4} k_i U_i$$

(where the U's have integer components) of $L = k_1$ for which $[L(U_1)]^2 + [L(U_2)]^2 + [L(U_3)]^2 + [L(U_4)]^2$ is a minimum.

We first make a few remarks about any general solution

$$V = \sum_{i=1}^{4} k_i W_i$$

of $L = k_1$ (where the W's have integer components). Putting $k_1 = 1$, $k_2 = k_3 = k_4 = 0$, we see that W_1 is a solution of $L = 1$. Also putting $k_1 = 0$, we see that

$$V = \sum_{i=2}^{4} k_i W_i$$

is a general solution in integers of $L = 0$. In other words, every lattice point on the hyperplane $L = 0$ is an I. L. C. of W_2, W_3, W_4 and conversely. Note that (408, −985, 0, 0), (780, 0, 985, 0), and (571, 0, 0, 985) are three obvious independent lattice points on $L = 0$. Hence W_2, W_3, and W_4 must be independent. Let X_2, X_3, X_4 be any G. C. F. of W_2, W_3, W_4. Then any I. L. C. of W_2, W_3, W_4 is an I. L. C. of X_2, X_3, X_4 and conversely, so that any lattice point on $L = 0$ is an I. L. C. of X_2, X_3, X_4 and conversely. So

$$V = \sum_{i=2}^{4} k_i X_i$$

is a general solution of $L = 0$. Now let X_1 be any solution of $L = 1$. Then

$$V = \sum_{i=1}^{4} k_i X_i$$

is a general solution of $L = k_1$, which we show as follows. Let V be any solution of $L = k_1$. Then $V - k_1 X_1$ is a solution of $L = 0$. So

$$V - k_1 X_1 = \sum_{i=2}^{4} k_i X_i .$$

By similar arguments, one can show conversely that if

$$V = \sum_{i=1}^{4} k_i X_i$$

is a general solution, then X_i is a solution of $L = 1$ and X_2, X_3, X_4 is a G. C. F. of W_2, W_3, W_4.

Now choose W_2, W_3, W_4 a minimal G. C. F. of V_2, V_3, V_4, namely

$$W_2 = (5, -2, 6, -1),$$
$$W_3 = (1, -6, -7, 7),$$
$$W_4 = (3, 13, 4, 9).$$

Also take W_1 a shortest vector of the form $V_1 - m_2V_2 - m_3V_3 - m_4V_4$ (see Construction 2), namely

$$W_1 = (0, 7, 0, 5).$$

By our method of choosing W_1, it is clearly a shortest solution of $L = 1$. By the previous discussion

$$V = \sum_{i=1}^{4} k_i W_i$$

is a general solution of $L = k_1$. We now undertake to show that it is the one which we are seeking. Let

$$V = \sum_{i=1}^{4} k_i U_i$$

be a solution with $\sum[L(U_i)]^2$ a minimum. By our previous discussion, W_2, W_3, W_4 is a G. C. F. of U_2, U_3, U_4, and hence is a minimal G. C. F. So by Theorem 13, $[L(W_2)]^2 + [L(W_3)]^2 + [L(W_4)]^2 \leqq [L(U_2)]^2 + [L(U_3)]^2 + [L(U_4)]^2$. Also U_1 is a solution of $L = 1$. As W_1 is a shortest such solution, $[L(W_1)]^2 \leqq [L(U_1)]^2$.

We pause to remark on the role of Construction 2 in the discovery of W_1. As indicated in the discussion following Construction 2, the shorter the vector one starts with, the less labor is involved in finding a minimum. So before applying Construction 2, we minimized V_1 by integral use of W_2. This gave W_1. Then we attempted to minimize W_1 by integral use of W_3 or W_4 or by unit use of combinations of W_2, W_3, W_4. When these attempts failed, then it was time to use Construction 2.

PROBLEM V. Let $L(x, y, z) = x \log 2 + y \log 3 + z \log 5$. Find the least positive value which $L(a, b, c)$ can assume subject to the restriction $a^2 + b^2 + c^2 \leqq 10000$, and find the a, b, and c which produce this value.

Clearly an alternative way of stating the problem is the following.

PROBLEM V. Find that lattice point distinct from the origin and lying within the sphere $x^2 + y^2 + z^2 = 10000$ which is closest to the plane $L(x, y, z) = 0$.

We proceed as follows. Take a linear form with integer coefficients, $K(x, y, z)$, such that the planes $L = 0$ and $K = 0$ nearly coincide. Then if there are lattice points on $K = 0$ within the sphere $x^2 + y^2 + z^2 = 10000$ (we shall henceforth refer to this sphere as σ), one of them is likely to be the point for which we are looking. So one restriction on the coefficients of K is that they should be small enough so that there are likely to be lattice points (besides the origin) on $K = 0$ and inside σ. This means, roughly speaking, that there should be lattice points on $K = 0$ whose components are numerically $\leqq 100$. So the coefficients of K should be numerically < 10000. For, let the coefficients of K each be numerically < 10000. Consider the three linear forms:

$$100\,K, \quad y, \quad z.$$

Their determinant, Δ, is numerically $< 1{,}000{,}000$, and so (by the theorem quoted just before Problem II) there are integer values of the variables, not

all zero, which make all three forms simultaneously $\leqq |\Delta|^{1/3} < 100$ in absolute value. As these integer values make K an integer, and as they make $|100K| < 100$, they must make $K = 0$. So there are integer values of x, y, and z, with $|y| < 100$, $|z| < 100$, which make $K = 0$. It does not follow, of course, that $|x| < 100$, but from the fact that $K = 0$ and $|y| < 100$ and $|z| < 100$, it follows that $|x|$ cannot be too large. So we wish a K with coefficients numerically < 10000 such that $K = 0$ and $L = 0$ shall coincide as nearly as possible. For this purpose we wish three integers k_1, k_2, k_3, all < 10000, which are nearly in the ratio $\log 2, \log 3, \log 5$. Hence

$$k_3 \log 2 - k_1 \log 5,$$

$$k_3 \log 3 - k_2 \log 5$$

must be small. As these are the components of $k_1V_1 + k_2V_2 + k_3V_3$, where

$$V_1 = (-\log 5, 0),$$

$$V_2 = (0, -\log 5),$$

$$V_3 = (\log 2, \log 3),$$

we search for short I. L. C.'s of V_1, V_2, V_3. This search can be carried out by applying Operation III to the three vectors (see Problem III). Using the values of k_1, k_2, k_3 which we get by use of Operation III, we take K to be

$$4296x + 6809y + 9975z.$$

Put $\alpha = ((4296)^2 + (6809)^2 + (9975)^2)^{1/2}$. We find that the angle, θ, between $K = 0$ and $L = 0$ is approximately $(0.00273/\alpha)$ radians. So at the surface of σ, the maximum separation between $K = 0$ and $L = 0$ is approximately 100θ, or $0.273/\alpha$. As the distance between the planes $K = 0$ and $K = 1$ is $1/\alpha$, we see that any point on $K = 0$ in σ must be closer to $L = 0$ than any point on $K = 1$ in σ. Likewise any point on $K = 1$ in σ must be closer to $L = 0$ than any point on $K = 2$ in σ. And so on. Note that any lattice point in space is on some one of the planes $K = m$ for an appropriately chosen m. For if (a, b, c) is the lattice point, choose $m = K(a, b, c)$. From all the above, we deduce the following procedure for solving our problem. First, find all lattice points on $K = 0$ in σ. If there are any (besides the origin), the closest to $L = 0$ is our answer. If there are no lattice points (except the origin) on $K = 0$ in σ, find all lattice points on $K = 1$ in σ. If there are any, the closest to $L = 0$ is our answer. If there are no lattice points on $K = 1$ in σ, find all lattice points on $K = 2$ in σ. And so on.

As a preliminary, we find a general solution in integers of $K = k_1$ for which the sum of the squares of the coefficients is a minimum (see Problem IV), namely

$$V = \sum_{i=1}^{3} k_i V_i,$$

where

$$V_1 = (54, -37, 2),$$
$$V_2 = (90, 15, -49),$$
$$V_3 = (71, -99, 37).$$

By the method which we used to find this solution, V_2 is a shortest non-zero vector with integer coefficients lying on $K = 0$. As $L(V_2) > 100$, there is no lattice point (except the origin) on $K = 0$ in σ. However V_1 is on $K = 1$ and is in σ. If we find all vectors of the form $V_1 - m_2V_2 - m_3V_3$ with length $\leqq 100$ (see Construction 1), we will have all lattice points on $K = 1$ in σ. The set of all lattice points on $K = 1$ and in σ consists of

$$(54, -37, 2),$$
$$(17, -62, 35),$$
$$(36, 52, -51).$$

Of these, the first is closest to $L = 0$, and we have

$$54 \log 2 - 37 \log 3 + 2 \log 5 = 0.000169$$

as the answer to our problem. If one is willing to increase the size of σ slightly in order to get a point much closer to $L = 0$, one would have

$$-90 \log 2 - 15 \log 3 + 49 \log 5 = 0.000027.$$

THEOREM 18. *Let U_1, V_1 and U_2, V_2 be minimal* G. C. F.*'s of $R_1, \cdots, R_m$. Then U_2 and V_2 must be two of $\pm U_1$, $\pm V_1$, $\pm U_1 \pm V_1$. Also if $L(U_1) < L(V_1)$, then U_2 is one of $\pm U_1$.*

Proof. Put $f(x, y) = [L(xU_1 + yV_1)]^2$. Since U_1, V_1 is minimal, we cannot perform Operation II on U_1, V_1, because if we could, we would get a U_3, V_3 with $L(U_3) + L(V_3) < L(U_1) + L(V_1)$, contradicting Theorem 12. Since we cannot perform Operation II on U_1, V_1, $f(x, y)$ satisfies the hypothesis of Theorem 6 (see the argument in the beginning of the proof of Theorem 9). Now U_2 is an I. L. C. of U_1 and V_1, say $U_2 = aU_1 + bV_1$. Then $f(1, 0) = [L(U_1)]^2 = [L(U_2)]^2 = [L(aU_1 + bV_1)]^2 = f(a, b)$. Note that the step $[L(U_1)]^2 = [L(U_2)]^2$ is based on Theorem 15. If $L(U_1) < L(V_1)$, then $f(1, 0) < f(0, 1)$. By Theorem 6, $f(a, b) \geqq f(0, 1)$ if $b \neq 0$, so if $L(U_1) < L(V_1)$, $b = 0$ and so $a = \pm 1$. If $L(U_1) = L(V_1)$, then $f(1, 0) = f(0, 1)$, so that we can have $b \neq 0$. However, we wish to show that in any case $|a| \leqq 1$ and $|b| \leqq 1$. Suppose $|a| > 1$ and $|b| > 1$. Then the α in Theorem 6 $\geqq 2$, contradicting $f(a, b) = f(0, 1)$. If either of $|a|$ or $|b|$ is zero and the other >1, then $\alpha \geqq 2$ again. If one of $|a|$ and $|b| > 1$ and the other $= 1$, then $|a| \neq |b|$ and neither is zero and $f(a, b) > f(0, 1)$ by Theorem 6. So $|a| \leqq 1$ and $|b| \leqq 1$. So U_2 must be one of the eight vectors mentioned. If we put $V_2 = cU_1 + dV_1$ and apply a similar argument, we get $|c| \leqq 1$ and $|d| \leqq 1$.

THEOREM 19. *Let U_1, V_1, W_1 and U_2, V_2, W_2 be minimal* G. C. F.*'s of $R_1, \cdots, R_m$. Then U_2, V_2, and W_2 must be* I. L. C.*'s of U_1, V_1, W_1 of the form $aU_1 + bV_1 + cW_1$, where $|a| \leqq 1$, $|b| \leqq 1$, and $|c| \leqq 1$. If $L(U_1) < L(V_1)$, then U_2 is one of $\pm U_1$. If $L(V_1) < L(W_1)$, then the allowable combinations for U_2 and V_2 cannot contain W_1.*

The proof is similar to that of Theorem 18. One point deserves mention. We start off by putting $f(x, y, z) = [L(xU_1 + yV_1 + zW_1)]^2$. We put $U_2 = aU_1 + bV_1 + cW_1$, so that $f(a, b, c) = f(1, 0, 0)$. If one of a, b, or c is zero, we proceed as follows. Suppose $a = 0$. Then think of $f(0, y, z)$ as a quadratic form $g(y, z)$ and note that $g(y, z)$ satisfies the hypothesis of Theorem 6. Then the argument of Theorem 18 will show that $|b| \leqq 1$ and $|c| \leqq 1$. The cases where none of a, b, or c is zero make use of Theorem 7.

THEOREM 20. *Let U_1, V_1, W_1, X_1 and U_2, V_2, W_2, X_2 be minimal* G. C. F.*'s of $R_1, \cdots, R_m$. Then U_2, V_2, W_2, and X_2 must be* I. L. C.*'s of U_1, V_1, W_1, X_1 of the form $aU_1 + bV_1 + cW_1 + dX_1$, where either all four of a, b, c, d have absolute value $\leqq 1$ or else exactly one of a, b, c, and d has absolute value 2 and the rest have absolute value 1. If $L(U_1) < L(V_1)$, then U_2 is one of $\pm U_1$. If $L(V_1) < L(W_1)$, then the allowable combinations for U_2 and V_2 cannot contain W_1 or X_1. If $L(W_1) < L(X_1)$, then the allowable combinations for U_2, V_2, and W_2 cannot contain X_1.*

Proof. The proof proceeds just like the proof of Theorem 19 up to the point where none of a, b, c, or d is zero and $f(a, b, c, d) = f(0, 0, 0, 1)$. If all of $|a|$, $|b|$, $|c|$, $|d|$ are $\geqq 2$, then in Theorem 8, $\beta \geqq 3$, and we have a contradiction. So let one of them be 1. Now if there are three different values among $|a|$, $|b|$, $|c|$, $|d|$, we have a contradiction again. Also if the two largest of $|a|$, $|b|$, $|c|$, $|d|$ are equal and $\geqq 2$, we can use the lemma of Theorem 8 to obtain a contradiction. So we are reduced to the case where three of $|a|$, $|b|$, $|c|$, $|d|$ equal 1 and the remaining one $\geqq 1$. We wish to show that the remaining one $\leqq 2$. The situation which we have corresponds to the situation that would result in Case 1 of Theorem 8 if we put $y = z = w = 1$. By going through the argument for this case, we see that $f(x, 1, 1, 1) > (0, 0, 0, 1)$ if $x > 2$.

To see a case where one of a, b, c, or d would have to be 2, take the usual definition of length and let the V's be

$$(2, 0, 0, 0),$$

$$(0, 2, 0, 0),$$

$$(0, 0, 2, 0),$$

$$(0, 0, 0, 2),$$

$$(1, 1, 1, 1).$$

Take the first three and the last to be U_1, V_1, W_1, X_1 and the last four to be U_2, V_2, W_2, X_2.

As an example of a case where there is a large number of minimal G. C. F.'s, take the usual definition of distance, and let

$$(1, 0, 0),$$
$$\left(\frac{1}{2}, \frac{\sqrt{3}}{2}, 0\right),$$
$$\left(\frac{1}{2}, -\frac{1}{2\sqrt{3}}, \frac{2}{\sqrt{6}}\right)$$

be a minimal G. C. F. of a set of R's. (The R's could consist of the minimal G. C. F. itself, of course.) Then if U_2, V_2, W_2 is another minimal G. C. F., $\pm U_2$ could be any one of

$$(1, 0, 0),$$
$$\left(\frac{1}{2}, \frac{\sqrt{3}}{2}, 0\right),$$
$$\left(-\frac{1}{2}, \frac{\sqrt{3}}{2}, 0\right),$$
$$\left(\frac{1}{2}, -\frac{1}{2\sqrt{3}}, \frac{2}{\sqrt{6}}\right),$$
$$\left(-\frac{1}{2}, -\frac{1}{2\sqrt{3}}, \frac{2}{\sqrt{6}}\right),$$
$$\left(0, \frac{1}{\sqrt{3}}, \frac{2}{\sqrt{6}}\right),$$

$\pm V_2$ could be any one of the remaining five, and $\pm W_2$ could be any one of three or four others. By trial, we find that sixteen distinct combinations of three out of the above six vectors will serve as a G. C. F. As $+$ or $-$ signs may be attached to three vectors in eight ways, we get a total of 8×16 G. C. F.'s. As these vectors are all the same length, they may be taken in any order to produce a minimal G. C. F., and as the order is a distinguishing property of minimal G. C. F.'s, we find that there are a total of $6 \times 8 \times 16$ or 768 possible minimal G. C. F.'s.

The purpose of these last three theorems is to enable one to find all minimal G. C. F.'s. For this, we proceed as follows. First, find a minimal G. C. F. by use of Operation II, III, or IV. Then the vectors of any other minimal G. C. F. must lie among the combinations listed in Theorems 18, 19, and 20. By trying all possible combinations, we can find all minimal G. C. F.'s. The information in Theorems 15, 16, and 17 is useful in eliminating unusable combinations. In this connection it is worth remarking that if U_1, V_1, W_1, X_1 is a minimal G. C. F. of $R_1, \cdots, R_m$ and U_2, V_2, W_2, X_2 is a G. C. F. of $R_1, \cdots, R_m$ and $L(U_1) = L(U_2)$, $L(V_1) = L(V_2)$, $L(W_1) = L(W_2)$, $L(X_1) = L(X_2)$, then U_2,

V_2, W_2, X_2 is a minimal G. C. F. of $R_1, \cdots, R_m$. For if U_2, V_2, W_2, X_2 is not minimal, then by Theorem 11, Operation IV can be performed, giving U_3, V_3, W_3, X_3 with $L(U_3) + L(V_3) + L(W_3) + L(X_3) < L(U_2) + L(V_2) + L(W_2) + L(X_2) = L(U_1) + L(V_1) + L(W_1) + L(X_1)$. This contradicts Theorem 14. Note that the result just stated does not hold without the condition that U_2, V_2, W_2, X_2 be a G. C. F. of $R_1, \cdots, R_m$. In fact, it is possible for U_2, V_2, W_2, X_2 to be such that U_2 is a shortest non-zero I. L. C. of $R_1, \cdots, R_m$, V_2 is a shortest I. L. C. of $R_1, \cdots, R_m$ which is independent of U_2, etc. without U_2, V_2, W_2, X_2 being a G. C. F. of $R_1, \cdots, R_m$. Under these circumstances U_2, V_2, W_2, X_2 naturally cannot be a minimal G. C. F. An illustration of this situation would be where

$$R_1 = (2, 0, 0, 0),$$
$$R_2 = (0, 2, 0, 0),$$
$$R_3 = (0, 0, 2, 0),$$
$$R_4 = (0, 0, 0, 2),$$
$$R_5 = (1, 1, 1, 1),$$

and $U_2 = R_1$, $V_2 = R_2$, $W_2 = R_3$, $X_2 = R_4$. The R's have a G. C. F., for instance, R_1, R_2, R_3, R_5, but U_2, V_2, W_2, X_2 is not a G. C. F., although all four of them are shortest non-zero I. L. C.'s of the R's.

We now consider some further matters related to Problem I. Let A, V_1, V_2, V_3, V_4 be as in Problem I, and let us find all minimal G. C. F.'s of V_1, V_2, V_3, V_4, basing our definition of length on A. We continue applying Operation IV until a minimal G. C. F. of V_1, V_2, V_3, V_4 is attained. Then we list all the combinations mentioned in Theorem 20. From these we eliminate a large number of unsuitable ones by use of Theorem 17. We get finally the following six vectors, the first three of length $\sqrt{2}$, and the last three of length 2.

$$R_1 = (1, 2, -1, 1),$$
$$R_2 = (-2, -6, 3, -3),$$
$$R_3 = (-1, -4, 2, -2),$$
$$R_4 = (0, 0, 1, -2),$$
$$R_5 = (0, -5, 2, -2),$$
$$R_6 = (0, -5, 3, -4).$$

If U, V, W, X is a minimal G. C. F. of V_1, V_2, V_3, V_4, then $\pm U$ and $\pm V$ can be any two of R_1, R_2, R_3 and $\pm W$, $\pm X$ can be any two of R_4, R_5, R_6. So there are a total of 576 possible minimal G. C. F.'s of V_1, V_2, V_3, V_4.

Now consider the problem of finding all minima of the quadratic form Q given in Problem I. $\pm R_1$, $\pm R_2$, and $\pm R_3$ each gives a minimum, but do they give all the minima? The fact that $\pm R_1$, $\pm R_2$, $\pm R_3$ are all the possibilities for U does not enable us to conclude that they are all shortest non-zero I. L. C.'s of V_1, V_2, V_3, V_4, because we have not proved the impossibility of having a

shortest non-zero I. L. C. of V_1, V_2, V_3, V_4 which is not a constituent of some minimal G. C. F. of V_1, V_2, V_3, V_4. However, if one observes the proof of Theorem 20 carefully, one will see that it is there proved that the combinations listed will include all shortest non-zero I. L. C.'s, although this is not stated in the theorem itself. So $\pm R_1$, $\pm R_2$, $\pm R_3$ are all of the minima of Q.

As we remarked while solving Problem I, A is the matrix of the quadratic form $2Q$. If we make a transformation of coordinates with matrix P, the new quadratic form will have matrix P^TAP.[6] If we wish the transformation to leave the number-theoretic properties of Q unchanged, P must have integral components and determinant ± 1. So we say that two symmetric positive definite matrices, A and B, are equivalent (in the number-theoretic sense) if there is a matrix P with integral components and determinant ± 1 such that $B = P^TAP$.

PROBLEM VI. With A as in Problem I and

$$B = \begin{Vmatrix} 34 & -3 & 36 & 92 \\ -3 & 2 & -4 & -10 \\ 36 & -4 & 44 & 110 \\ 92 & -10 & 110 & 276 \end{Vmatrix},$$

determine if A and B are equivalent, and if they are, find P.

First, a few general considerations. Suppose A and B are equivalent, say $B = P^TAP$. Let V be any vector. Then the length of V relative to B, $L_B(V)$, is equal to $(V^TBV)^{1/2} = ((PV)^TA(PV))^{1/2} = L_A(PV)$. So if U is a shortest non-zero I. L. C. of $R_1, \cdots, R_m$ relative to B, then PU is a shortest non-zero I. L. C. of $PR_1, \cdots, PR_m$ relative to A. Let

$$S_1 = (1, 0, 0, 0),$$
$$S_2 = (0, 1, 0, 0),$$
$$S_3 = (0, 0, 1, 0),$$
$$S_4 = (0, 0, 0, 1),$$

and let U, V, W, X be a minimal G. C. F. of S_1, S_2, S_3, S_4 relative to B, for instance

$$U = (0, 1, 0, 0),$$
$$V = (1, 0, 4, -2),$$
$$W = (0, 0, 5, -2),$$
$$X = (-2, 0, -6, 3).$$

Then PU, PV, PW, PX is a minimal G. C. F. of PS_1, PS_2, PS_3, PS_4 relative to A. However, since the components of P are integers, the components of PS_1, PS_2, PS_3, PS_4 are also integers, and so each of PS_1, PS_2, PS_3, PS_4 is an I. L. C. of S_1, S_2, S_3, S_4. Conversely, any vector T with integer com-

[6] M. Bôcher, *Introduction to Higher Algebra*, p. 129, Theorem 1.

ponents is an I. L. C. of PS_1, PS_2, PS_3, PS_4, because if we try to solve $aPS_1 + bPS_2 + cPS_3 + dPS_4 = T$ for a, b, c, and d, we find that we have four linear equations to be solved whose matrix is P, and so whose determinant is ± 1. Hence the equations can be solved in integers. So PS_1, PS_2, PS_3, PS_4 is a G. C. F. of S_1, S_2, S_3, S_4. So any minimal G. C. F. of PS_1, PS_2, PS_3, PS_4, for instance PU, PV, PW, PX, is a minimal G. C. F. of S_1, S_2, S_3, S_4. So one of the minimal G. C. F.'s of S_1, S_2, S_3, S_4 relative to A (all of which we found a while back) must be PU, PV, PW, PX. If we can guess the right minimal G. C. F. relative to A, we can solve for P. Just to get going, let us conjecture that perhaps R_1, R_2, R_4, R_5 is PU, PV, PW, PX. Before solving for P, it is well to test the reasonableness of this conjecture. We find that $L_A(R_1) = L_B(U)$, $L_A(R_2) = L_B(V)$, $L_A(R_4) = L_B(W)$, $L_A(R_5) = L_B(X)$. This is as it should be. However $U^TBV = (PU)^TA(PV)$. So we should have $U^TBV = R_1^TAR_2$. As a matter of fact, we have instead $U^TBV = -R_1^TAR_2$. This suggests replacing R_2 by $-R_2$. For a similar reason we replace R_5 by $-R_5$. Now we can find no reason why we should not have $R_1 = PU$, $-R_2 = PV$, $R_4 = PW$, and $-R_5 = PX$, so we solve these for P, getting

$$P = \begin{Vmatrix} -6 & 1 & -8 & -20 \\ -28 & 2 & -34 & -85 \\ 13 & -1 & 17 & 42 \\ -13 & 1 & -18 & -44 \end{Vmatrix}.$$

With this value of P, we have $|P| = 1$ and $B = P^TAP$. So A and B are equivalent.

The reader may wonder whether it was just luck that we found P so readily. Clearly, with 576 possible minimal G. C. F.'s relative to A to choose from, if only one would have given P, then we certainly were lucky. However, as a matter of fact, there are quite a large number of P's such that $B = P^TAP$, and had we tried some other minimal G. C. F. relative to A, we would have merely found some other P. The fact is that there are quite a number of P's with integral components and determinant ± 1 such that $P^TAP = A$. Such a P we call an automorphism (in the number-theoretic sense) of A. By using two different properly chosen minimal G. C. F.'s relative to A, we can find automorphisms of A. For instance, if we choose P so that $PR_3 = R_1$, $PR_2 = -R_2$, $PR_5 = -R_4$, and $PR_4 = -R_5$, we get

$$P = \begin{Vmatrix} -1 & 0 & 0 & 0 \\ -6 & -4 & -25 & -15 \\ 3 & 3 & 16 & 9 \\ -3 & -4 & -20 & -11 \end{Vmatrix}.$$

Then $|P| = 1$ and $A = P^TAP$, so that P is an automorphism of A.

As any automorphism of A carries a minimal G. C. F. into a different minimal

G. C. F., and as there are only a finite number of distinct minimal G. C. F.'s, there can be only a finite number of automorphisms of A. As the automorphisms of A form a group, it follows that any automorphism of A must be of finite order. For the P given above, $P^2 = 1$. In order to count the number of automorphisms of A, note that if Y_1, Y_2, Y_3, Y_4 and Z_1, Z_2, Z_3, Z_4 are two minimal G. C. F.'s such that $Y_i^T A Y_j = Z_i^T A Z_j$ for all i and j, then there must be an automorphism P such that $Y_i = PZ_i$. To show this, first choose G and H so that $GS_i = Y_i$ and $HS_i = Z_i$. Then since Y_1, Y_2, Y_3, Y_4 is a G. C. F. of S_1, S_2, S_3, S_4, it follows that G has integral components and determinant ± 1. Similarly for H. Now $Y_i^T A Y_j = S_i^T (G^T A G) S_j =$ the element in the i-th row and j-th column of $G^T A G$. Similarly $Z_i^T A Z_j = S_i^T (H^T A H) S_j =$ the element in the i-th row and j-th column of $H^T A H$. So, since $Y_i^T A Y_j = Z_i^T A Z_j$, the matrices $G^T A G$ and $H^T A H$ are identical. From $G^T A G = H^T A H$, we clearly get $(GH^{-1})^T A (GH^{-1}) = A$. So GH^{-1} is an automorphism of A. However $(GH^{-1})Z_i = (GH^{-1})(HS_i) = GS_i = Y_i$. To facilitate choosing such sets of Y's and Z's, we list the values of $R_i^T A R_j$ in Table IV at the end of the paper. We might as well choose the same Y's every time, say $Y_1 = R_1, Y_2 = R_2, Y_3 = R_4, Y_4 = R_5$. Then Z_1 can be any one of $\pm R_1$, $\pm R_2$, $\pm R_3$. Whichever of these we choose for Z_1, if any of the remaining five be chosen for Z_2, it will give $Z_2^T A Z_2$ the right value, but only two can be chosen which will give $Z_1^T A Z_2$ the right value. So there are twelve possible choices for Z_1, Z_2. Similarly Z_3 can be any one of $\pm R_4, \pm R_5, \pm R_6$, with two choices left for Z_4. So there are 144 different automorphisms of A. This count takes P and $-P$ as distinct automorphisms, so that there are 72 essentially distinct automorphisms of A.

Prof. R. J. Walker has made an interesting suggestion. In several of the problems which we considered, we are primarily interested in obtaining I. L. C.'s with small components, and we were able to solve the problems by use of our algorithm only because the short vectors which are obtained by use of it necessarily have small components. This suggests that we define the length of a vector to be the absolute value of its numerically largest component. This would give quite a satisfactory metric. Furthermore, with this metric the problem of finding a shortest vector of the form $U - nV$ could be easily solved. Hence we could define Operations II, III, and IV as above, only with this new metric in mind. The particular cause which is responsible for the failure of the algorithms of this paper in the case of five or more dimensions would fail to apply to the new metric. Hence, with the new metric there is at least a possibility that there always exists a minimal G. C. F. in any number of dimensions, and that a generalization of Operations II, III, and IV (perhaps with slight modifications) will suffice to find a minimal G. C. F. However, as yet we cannot prove anything useful about the new metric in more than two dimensions.

CORNELL UNIVERSITY.

TABLE I

	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_{11}	V_{12}	V_{13}	V_{14}	V_{15}
$(1, 0, 0, 0) = V_1$	14	18	90	45	4										
$(0, 1, 0, 0) = V_2$	18	28	134	66											
$(0, 0, 1, 0) = V_3$	90	134	652	324	44	46									
$(0, 0, 0, 1) = V_4$	45	66	324	162	21	24	177								
$(-1, 1, 0, 0) = V_2 - V_1 = V_5$	4		44	21	6	−2	2	3	12	0	−2	3	7	1	1
$(2, -1, 0, 0) = V_1 - V_5 = V_6$			46	24	−2	12	60	30	0	4					
$(7, -7, 1, 0) = V_3 - 7V_5 = V_7$				177	2	60	330	171							
$(3, -3, 0, 1) = V_4 - 3V_5 = V_8$					3	30	171	90	21	15	15				
$(-3, -2, 1, 0) = V_7 - 5V_6 = V_9$					12	0		21	30						
$(-1, -4, 1, 0) = V_9 - 2V_5 = V_{10}$					0	4		15		6	−2	3	7	7	1
$(3, 3, -1, 0) = V_6 - V_{10} = V_{11}$					−2			15		−2	10	19	−1	1	3
$(5, 5, -2, 1) = V_8 - 2V_{10} = V_{12}$					3					3	19	54			
$(-1, -1, 0, 1) = V_{12} - 2V_{11} = V_{13}$					7					7	−1		18		
$(0, -2, 0, 1) = V_{13} - V_5 = V_{14}$					1					7	1			10	
$(1, 2, -1, 1) = V_{14} - V_{10} = V_{15}$					1					1	3				2

TABLE II

x	y	z	w	L_1	L_2	L_3	L_4
−2557879	2869678	3909845	−9976556	0	1	2	1
−4248882	4766810	6494627	−16572015	−1	3	−2	0
−3586719	4023931	5482478	−13989365	−3	−1	1	−2
−6144598	6893609	9392323	−23965921	−3	0	3	−1
−1028840	1154253	1572633	−4012809	−3	−2	−1	−3

TABLE III

a	b	c	
1	3	1	$1 \leqq z \leqq 5$
2	6	1	$2 \leqq z \leqq 12$
7	22	4	$7 \leqq z \leqq 600$
466	1464	269	$466 \leqq z \leqq 3100$
1010	3173	583	$1010 \leqq z \leqq 6600$
2493	7832	1439	$2493 \leqq z \leqq 11{,}000$
8023	25205	4631	$8023 \leqq z \leqq 210{,}000$
62701	196981	36192	$62701 \leqq z \leqq 1{,}000{,}000$

TABLE IV

	R_1	R_2	R_3	R_4	R_5	R_6
R_1	2	−1	1	0	0	0
R_2	−1	2	1	0	0	0
R_3	1	1	2	0	0	0
R_4	0	0	0	4	−2	2
R_5	0	0	0	−2	4	2
R_6	0	0	0	2	2	4

POSITIVE DEFINITE FUNCTIONS ON SPHERES

By I. J. Schoenberg

1. **Introduction.** Let S_2 denote the ordinary spherical shell of radius one and center o and let $p_1, p_2, \cdots, p_n$ be n arbitrary points of S_2. Let p_ip_k denote the spherical distance between the points p_i, p_k. For n real variables $x_1, x_2, \cdots, x_n$ we have the following inequality

$$|\sum \overrightarrow{op_i} \cdot x_i|^2 = \sum_{i,k=1}^{n} \cos(p_ip_k)\, x_i x_k \geqq 0, \tag{1.1}$$

which is equivalent to the determinant inequality

$$\det |\cos(p_ip_k)|_{1,n} \geqq 0$$

for arbitrary points p_i and arbitrary n. This property of the function $g(t) = \cos t$ in relation to the space S_2 is expressed by saying that $\cos t$ is positive definite in S_2. The general definition is as follows. Let M be a metric space with the distance function pq. A real continuous function $g(t)$ $(0 \leqq t \leqq \text{diameter of } M)$ is said to be positive definite (p. d.) in M if we have

$$\sum_{i,k=1}^{n} g(p_ip_k)\, x_i x_k \geqq 0, \tag{1.2}$$

for any n points $p_1, \cdots, p_n$ of M, arbitrary real x_i, and all $n = 2, 3, \cdots$.

We denote this class of functions by the symbol $\mathfrak{P}(M)$. It enjoys the following useful closure properties:[1]

I. If $g_1(t) \in \mathfrak{P}(M)$, $g_2(t) \in \mathfrak{P}(M)$, also $c_1g_1(t) + c_2g_2(t) \in \mathfrak{P}(M)$, provided $c_1 \geqq 0$, $c_2 \geqq 0$.

II. The same assumptions imply also that $g_1(t)g_2(t) \in \mathfrak{P}(M)$.

III. If $g_n(t) \in \mathfrak{P}(M)$, $g_n(t) \to g(t)$ as $n \to \infty$, and $g(t)$ is continuous, then also $g(t) \in \mathfrak{P}(M)$.

In the present note we are concerned with the classes $\mathfrak{P}(S_m)$ and $\mathfrak{P}(S_\infty)$ corresponding to the unit spheres in the Euclidean space E_{m+1} and the Hilbert space H respectively.

Returning to S_2, we have noticed that $\cos t \in \mathfrak{P}(S_2)$. It will be shown below that also $P_n(\cos t) \in \mathfrak{P}(S_2)$, where P_n is a Legendre polynomial. By the above mentioned closure properties it is now apparent that also

$$g(t) = \sum_{n=0}^{\infty} a_n P_n(\cos t) \in \mathfrak{P}(S_2), \tag{1.3}$$

provided $a_n \geqq 0$ $(n = 0, 1, \cdots)$ and $\sum a_n$ converges. This formula will be shown to furnish the most general element of $\mathfrak{P}(S_2)$.

Received August 25, 1941; presented to the American Mathematical Society, September 11, 1940.

[1] See [6], §2. Numbers in brackets refer to the bibliography at the end of the paper.

Let $P_n^{(\lambda)}(\cos t)$ be the ultraspherical polynomials defined by the expansion

$$(1.4) \qquad (1 - 2r\cos t + r^2)^{-\lambda} = \sum_{n=0}^{\infty} r^n P_n^{(\lambda)}(\cos t), \qquad (\lambda > 0).$$

For $\lambda = 0$ we set

$$P_n^{(0)}(\cos t) = \cos nt = T_n(\cos t).$$

Our result (1.3) extends to S_m $(m = 1, 2, \cdots)$ as follows. The most general element of $\mathfrak{P}(S_m)$ is

$$(1.5) \qquad g(t) = \sum_{n=0}^{\infty} a_n P_n^{(\lambda)}(\cos t), \qquad (\lambda = \tfrac{1}{2}(m-1)),$$

provided all $a_n \geqq 0$ and that we have convergence for $t = 0$. This will follow readily from classical properties of ultraspherical polynomials (§3).

An obvious property of the class $\mathfrak{P}(M)$ is as follows. If $M \subset N$, then $\mathfrak{P}(M) \supset \mathfrak{P}(N)$. As we may assume

$$S_1 \subset S_2 \subset \cdots \subset S_m \subset \cdots \subset S_\infty,$$

it follows that

$$\mathfrak{P}(S_1) \supset \mathfrak{P}(S_2) \supset \cdots \supset \mathfrak{P}(S_m) \supset \cdots \supset \mathfrak{P}(S_\infty).$$

In fact, $\mathfrak{P}(S_\infty)$ is identical with the intersection of all classes $\mathfrak{P}(S_m)$ $(m = 1, 2, \cdots)$. Since $\cos t \in \mathfrak{P}(S_m)$ for all m, we have $\cos t \in \mathfrak{P}(S_\infty)$ and, by the closure property II, also $(\cos t)^n \in \mathfrak{P}(S_\infty)$. The closure properties I and III show that

$$(1.6) \qquad g(t) = \sum_{n=0}^{\infty} a_n(\cos t)^n \in \mathfrak{P}(S_\infty),$$

provided $a_n \geqq 0$ and $\sum a_n$ converges. It will be shown that the functions $g(t)$ of the form (1.6) exhaust the class $\mathfrak{P}(S_\infty)$(§4).[2]

[2] This result establishes the *converse* of Problem 37 of Pólya and Szegö, [5], p. 107. This converse may be stated as follows. *Let* $F(x)$, $(-1 \leqq x \leqq 1)$, *be a continuous function enjoying the following property. If the quadratic form* $\sum_1^n a_{ik}x_ix_k$, $(-1 \leqq a_{ik} \leqq 1; i = 1, \cdots, n; n = 2, 3, \cdots)$, *is positive, the form* $\sum_1^n F(a_{ik})x_ix_k$ *should also be positive. Then* $F(x)$ *is necessarily of the form*

$$F(x) = \sum_0^\infty a_\nu x^\nu, \qquad (a_n \geqq 0, -1 \leqq x \leqq 1).$$

This result is implied by the fact that every element of $\mathfrak{P}(S_\infty)$ is of the form (1.6). Indeed, it suffices to assume $a_{11} = a_{22} = \cdots = a_{nn} = 1$ and to notice (i) that then $a_{ik} = \cos(p_ip_k)$, where $p_1, \cdots, p_n$ are n appropriate points of S_{n-1}, (ii) that the assumptions of our proposition imply that $F(\cos t)$ is positive definite in S_n.

A second application of our results is as follows. The classical expansion of $(\cos t)^n$ in terms of Legendre polynomials has *positive* coefficients. The reason for this is now obvious in view of the relations $(\cos t)^n \in \mathfrak{P}(S_\infty) \subset \mathfrak{P}(S_2)$ and formula (1.3). We may predict in the same way that the expansion of $P_n^{(\lambda)}(\cos t)$ in terms of the polynomials $P_m^{(\mu)}(\cos t)$, $(m = 0, 1, \cdots; 0 \leqq \mu < \lambda)$, will likewise have positive coefficients, provided 2μ, 2λ are integers.

The fifth and last section of this paper is devoted to problems of isometric imbedding in Hilbert space. We determine all metric transforms of the sphere T_∞ which can be imbedded isometrically in H.

2. **Some classical results concerning ultraspherical developments.** We return to the spherical shell S_m, of E_{m+1}, defined by the equation $x^2 + x_1^2 + \cdots + x_m^2 = 1$. If p is an arbitrary point of S_m and $0 \leqq r < 1$, let p_r be that point on the radius $\overline{op}$ such that $op_r = r$. Furthermore, let $F(p)$ be a real and continuous point function defined in S_m and let[3]

$$F(p_r) = \frac{1}{\omega_m} \int_{S_m} \frac{1 - r^2}{(1 - 2r \cos pp' + r^2)^{\frac{1}{2}(m+1)}} F(p')\, d\omega_{p'}, \tag{2.1}$$

where $\omega_m = 2\pi^{\frac{1}{2}(m+1)}\Gamma^{-1}[\frac{1}{2}(m+1)]$ is the "area" of S_m and pp' denotes a spherical distance. For $m = 2$, (2.1) reduces to the ordinary Poisson integral which solves the boundary value problem for harmonic functions inside the sphere. A classical result concerning the Poisson integral (2.1) is the limiting relation

$$\lim_{r \to 1} F(p_r) = F(p). \tag{2.2}$$

Differentiating (1.4) with respect to r and setting $t = pp'$, we get

$$\frac{1 - r^2}{(1 - 2r \cos pp' + r^2)^{\lambda+1}} = \sum_{n=0}^{\infty} \left(\frac{n}{\lambda} + 1\right) r^n P_n^{(\lambda)}(\cos pp'), \quad (0 \leqq r < 1). \tag{2.3}$$

Let $\lambda = \frac{1}{2}(m - 1)$. Multiplying (2.3) by $F(p')\, d\omega_{p'}/\omega_m$ and integrating both sides over S_m we now get, in view of (2.1), the development[4]

$$F(p_r) = \sum_{n=0}^{\infty} \frac{n + \lambda}{\lambda \omega_m} r^n \int_{S_m} F(p') P_n^{(\lambda)}(\cos pp')\, d\omega_{p'}, \qquad (\lambda = \tfrac{1}{2}(m - 1)). \tag{2.4}$$

The ultraspherical development of $F(p)$ is now given by

$$F(p) \sim \sum_{n=0}^{\infty} \frac{n + \lambda}{\lambda \omega_m} \int_{S_m} F(p') P_n^{(\lambda)}(\cos pp')\, d\omega_{p'}. \tag{2.5}$$

Our relation (2.2) expresses the well-known fact that the ultraspherical expansion (2.5) is Abel-summable at every point p of S_m to the sum $F(p)$.

We shall now derive the special form of the expansion (2.5) for the case of a

[3] See [1], p. 198.

[4] See [1], formula (18), p. 207.

zonal function $F(p)$. For this purpose we express the cooordinates of the point $p = (x, x_1, \cdots, x_m)$ in terms of polar coordinates as follows:

$$
\begin{aligned}
x &= \cos\theta, \\
x_1 &= \sin\theta\cos\theta_1, \\
x_2 &= \sin\theta\sin\theta_1\cos\theta_2, \\
&\cdots\cdots \\
x_{m-2} &= \sin\theta\sin\theta_1\cdots\cos\theta_{m-2}, \\
x_{m-1} &= \sin\theta\sin\theta_1\cdots\sin\theta_{m-2}\cos\phi, \qquad (0 \leqq \theta, \leqq \pi, 0 \leqq \phi \leqq 2\pi), \\
x_m &= \sin\theta\sin\theta_1\cdots\sin\theta_{m-2}\sin\phi, \qquad (\nu = 0, 1, \cdots, m-2),
\end{aligned} \tag{2.6}
$$

and consider the special case

$$F(p) = f(\cos\theta).$$

The integral on the right hand side of (2.5) now becomes

$$
\begin{aligned}
I_n = \int_0^{\pi}\cdots\int\int_0^{2\pi} & f(\cos\theta')P_n^{(\lambda)}(\cos pp') \\
& \cdot\sin^{m-1}\theta'\sin^{m-2}\theta_1'\cdots\sin\theta_{m-2}'\,d\theta'\cdots d\theta_{m-2}'\,d\phi'.
\end{aligned} \tag{2.7}
$$

Integrating first with respect to the variables $\theta_1', \cdots, \theta_{m-2}', \phi'$, we get

$$I_n = \int_0^{\pi} f(\cos\theta')J_n\sin^{m-1}\theta'\,d\theta', \tag{2.8}$$

where we set

$$J_n = \int_0^{\pi}\cdots\int\int_0^{2\pi} P_n^{(\lambda)}(\cos pp')\sin^{m-2}\theta_1'\cdots\sin\theta_{m-2}'\,d\theta_1'\cdots d\theta_{m-2}'\,d\phi'. \tag{2.9}$$

This integral is now readily reducible to a simple integral as follows. Consider the two points p_1 and p_1' of polar coordinates

$$p_1 = (\tfrac{1}{2}\pi, \theta_1, \cdots, \theta_{m-2}, \phi), \qquad p_1' = (\tfrac{1}{2}\pi, \theta_1', \cdots, \theta_{m-2}', \phi'),$$

both lying in the unit sphere S_{m-1} defined by $\theta = \frac{1}{2}\pi$. Obviously

$$\cos pp' = \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos p_1p_1'. \tag{2.10}$$

Now we notice that J_n as given by (2.9) amounts to an integration of $P_n^{(\lambda)}(\cos pp')$ over S_{m-1}. If we take in S_{m-1} a new system of polar coordinates ($\zeta = p_1p_1'$, $\zeta_1, \cdots, \zeta_{m-3}, \psi$), of pole p_1, we obtain

$$J_n = \int_0^{\pi}\cdots\int\int_0^{2\pi} P_n^{(\lambda)}(\cos pp')\sin^{m-2}\zeta\sin^{m-3}\zeta_1\cdots\sin\zeta_{m-3}\,d\zeta\cdots d\zeta_{m-3}\,d\psi.$$

Since $\cos pp'$ depends only on ζ, as shown by (2.10), the remaining integrations may be carried out leading to the expression

$$J_n = \omega_{m-2} \int_0^\pi P_n^{(\lambda)}(\cos pp') \sin^{m-2} \zeta \, d\zeta.$$

Now, since[5]

$$\int_0^\pi P_n^{(\lambda)}(\cos pp') \sin^{2\lambda-1} \zeta \, d\zeta = \frac{\Gamma(\lambda)\Gamma(\frac{1}{2})\Gamma(n+1)\Gamma(2\lambda)}{\Gamma(\lambda+\frac{1}{2})\Gamma(n+2\lambda)} P_n^{(\lambda)}(\cos\theta) P_n^{(\lambda)}(\cos\theta'),$$

J_n is explicitly computed. Substituting its value into (2.8) we find that the general expansion (2.5) reduces to[6]

$$f(\cos\theta) \sim \sum_{n=0}^{\infty} \frac{(n+\lambda)\Gamma(\lambda)}{\Gamma(\lambda+\frac{1}{2})\Gamma(\frac{1}{2})} \frac{\Gamma(n+1)\Gamma(2\lambda)}{\Gamma(n+2\lambda)} P_n^{(\lambda)}(\cos\theta)$$

(2.11)

$$\cdot \int_0^\pi P_n^{(\lambda)}(\cos\theta') f(\cos\theta') \sin^{m-1}\theta' \, d\theta'.$$

This expansion is, of course, also Abel-summable. It will be applied in the next section to functions $f(\cos\theta)$ which are positive definite in S_m.

3. **On positive definite functions in S_m.** Let $f(x)$ be real and continuous in the interval $-1 \leqq x \leqq 1$ and such that $f(\cos t)$ is p. d. in S_m. As stated in the introduction this amounts to the inequality

$$\sum_1^n f(\cos p_i p_k) x_i x_k \geqq 0, \qquad (p_i \in S_m,\ x_i \text{ real}). \tag{3.1}$$

According to W. H. Young this quadratic form requirement is equivalent to the integral inequality[7]

$$I(h) = \int_{S_m} \int_{S_m} f(\cos pp') h(p) h(p') \, d\omega_p \, d\omega_{p'} \geqq 0 \tag{3.2}$$

for an arbitrary continuous point function $h(p)$ in S_m.

A consequence of this re-statement is as follows. If $h(p) \equiv 1$, (3.2) becomes

$$I(1) = \int_{S_m} \left\{ \int_{S_m} f(\cos pp') \, d\omega_{p'} \right\} d\omega_p = \omega_m \int_{S_m} f(\cos pp') \, d\omega_{p'}$$

since the last integral is clearly independent of p. Since $I(1) \geqq 0$, we have proved that

$$\int_{S_m} f(\cos pp') \, d\omega_{p'} \geqq 0, \tag{3.3}$$

provided $f(\cos t)$ is p. d. in S_m.

[5] See [4], formula (1), p. 203.

[6] See [4], p. 198. We need below the fact that the expansion (2.11) is a special case of (2.5). Our derivation of (2.11) from (2.5) is well known as an analogue of a classical reduction. However, being unable to find it in the literature, I developed it here in detail.

[7] See [8].

A further result which we need is as follows. *The ultraspherical polynomials*

$$P_n^{(\lambda)}(\cos t), \quad (n = 0, 1, \cdots ; \lambda = \tfrac{1}{2}(m-1)),$$

are all p. d. in S_m.

This statement is trivial for $m = 1$ ($\lambda = 0$) in view of the cosine addition formula. Let $m \geqq 2$. In order to prove that $P_n^{(\lambda)}(\cos t)$ is p. d. in S_m we proceed by induction assuming $P_n^{(\lambda-\frac{1}{2})}(\cos t)$ to be p. d. in S_{m-1}. Suppose $p_i \in S_m$ ($i = 1, \cdots, N$) and associate with the point p_i a point p_i', on the "equator" S_{m-1} of equation $\theta = \frac{1}{2}\pi$, such that the last $m-1$ polar coordinates $\theta_1, \cdots, \phi$ of both points p_i and p_i' agree. As remarked before we have

$$\cos p_i p_k = \cos \theta^i \cos \theta^k + \sin \theta^i \sin \theta^k \cos p_i' p_k' .$$

By the addition formula for ultraspherical polynomials we may write[8]

$$P_n^{(\lambda)}(\cos p_i p_k) = \sum_{s=0}^{n} c_{n,\lambda,s} P_n^{\lambda,s}(\cos \theta^i) P_n^{\lambda,s}(\cos \theta^k) P_n^{(\lambda-\frac{1}{2})}(\cos p_i' p_k'),$$

where $P_n^{\lambda,s}$ are the real polynomials associated to $P_n^{(\lambda)}$ and $c_{n,\lambda,s}$ are *positive* coefficients whose values are here of no concern. But then

$$(3.4) \qquad \sum_{i,k=1}^{N} P_n^{(\lambda)}(\cos p_i p_k)\xi_i \xi_k = \sum_{s=0}^{n} c_{n,\lambda,s} \sum_{i,k=1}^{N} P_n^{(\lambda-\frac{1}{2})}(\cos p_i' p_k')\eta_i \eta_k \geqq 0,$$

where $\eta_i = P_n^{\lambda,s}(\cos \theta^i)\xi_i$. The expression (3.4) is indeed non-negative since $P_n^{(\lambda-\frac{1}{2})}(\cos t)$ was assumed to be p. d. in S_{m-1}. This completes our proof.

We may now establish the following theorem.[9]

THEOREM 1. *A necessary and sufficient condition in order that* $f(\cos \theta)$ *be positive definite in* S_m *is that the ultraspherical expansion* (2.11) *have non-negative coefficients in which case the series* (2.11) *converges throughout* $0 \leqq \theta \leqq \pi$ *absolutely and uniformly to the sum* $f(\cos \theta)$. *The most general* $f(\cos \theta)$ *which is p. d. in* S_m *is therefore given by the expansion*

$$(3.5) \qquad f(\cos \theta) = \sum_{n=0}^{\infty} a_n P_n^{(\lambda)}(\cos \theta), \qquad (a_n \geqq 0, \lambda = \tfrac{1}{2}(m-1)),$$

provided the series converges for $\theta = 0$.

Indeed, let $f(\cos \theta)$ be p. d. in S_m. The coefficient of $P_n^{(\lambda)}(\cos \theta)$ in (2.11) may be written as

$$\int_0^{\pi} P_n^{(\lambda)}(\cos \theta') f(\cos \theta') \sin^{m-1} \theta' \, d\theta' = \frac{1}{\omega_{m-1}} \int_{S_m} P_n^{(\lambda)}(\cos ap') f(\cos ap') \, d\omega_{p'},$$

where a is the point of S_m of coordinates $x = 1$, $x_1 = \cdots = x_m = 0$. But then the last integral is visibly positive for the following reason. Since $P_n^{(\lambda)}(\cos t)$ and $f(\cos t)$ are both p. d. in S_m, their product also enjoys this

[8] See [4], formula (6), p. 182.

[9] Recently Bochner has extended this theorem by means of the theory of generalized spherical harmonics of E. Cartan and H. Weyl. See [2], §§III and IV.

property. Now (3.3) shows that all coefficients of (2.11) are non-negative. We may therefore re-write (2.11) in the form

$$f(\cos\theta) \sim \sum_{n=0}^{\infty} a_n P_n^{(\lambda)}(\cos\theta), \qquad (a_n \geqq 0). \tag{3.6}$$

On the other hand, we know this series to be Abel-summable for all θ, hence, in particular, for $\theta = 0$. Thus

$$\sum_{n=0}^{k} a_n \mid P_n^{(\lambda)}(\cos\theta) \mid \leqq \sum_{n=0}^{k} a_n P_n^{(\lambda)}(1) \leqq \lim_{r\to 1} \sum_{0}^{\infty} a_n r^n P_n^{(\lambda)}(1) = f(1).$$

This shows that the series (3.6) is absolutely and uniformly convergent for all θ $(0 \leqq \theta \leqq \pi)$, hence convergent to its Abel-sum which is $f(\cos\theta)$.

The converse to the effect that the convergent series (3.5) defines a p. d. $f(\cos\theta)$ is clear. Indeed, $f(\cos\theta)$ is obviously continuous because the series (3.5) must converge uniformly. As $f(\cos\theta)$ thus appears as the continuous limit of a sequence of p. d. functions, it is p. d. itself.

4. **On positive definite functions in S_∞.** If $f(\cos\theta)$ is p. d. in S_∞ it is also p. d. in S_m. By Theorem 1 we are therefore assured to have an expansion with non-negative coefficients

$$f(\cos\theta) = \sum_{n=0}^{\infty} a_n(\lambda) P_n^{(\lambda)}(\cos\theta), \qquad (a_n(\lambda) \geqq 0,\ 0 \leqq \theta \leqq \pi), \tag{4.1}$$

which is valid for all values of λ of the form $\lambda = \frac{1}{2}(m-1)$, $(m = 1, 2, 3, \cdots)$. Setting

$$p_n^{\lambda}(\cos\theta) = \frac{P_n^{(\lambda)}(\cos\theta)}{P_n^{(\lambda)}(1)}, \tag{4.2}$$

we have a similar expansion

$$f(\cos\theta) = \sum_{n=0}^{\infty} b_n(\lambda) p_n^{\lambda}(\cos\theta), \qquad (b_n(\lambda) \geqq 0,\ 0 \leqq \theta \leqq \pi). \tag{4.3}$$

Since (see [4], p. 95)

$$\lim_{\lambda\to\infty} p_n^{\lambda}(\cos\theta) = \cos^n\theta, \tag{4.4}$$

we should expect to derive from (4.3), by letting $\lambda \to \infty$, the following theorem.

THEOREM 2. *A function $f(\cos\theta)$ which is positive definite in S_∞ is necessarily of the form*

$$f(\cos\theta) = \sum_{n=0}^{\infty} a_n \cos^n\theta, \qquad (a_n \geqq 0). \tag{4.5}$$

Now (4.4) obviously holds uniformly in θ. This fact, however, will not suffice to prove Theorem 2 due to our scant information concerning the coefficients

$b_n(\lambda)$ of (4.3). What we shall actually need is that, for a fixed value of θ, (4.4) holds uniformly for all n. This we state as a separate lemma.

LEMMA 1. *Let θ be such that $0 < \theta < \pi$. If $\epsilon > 0$ is arbitrarily small, we have*

$$|\, p_n^{\lambda}(\cos\theta) - \cos^n\theta \,| < \epsilon \qquad \text{for all } n = 0, 1, 2, \cdots, \tag{4.6}$$

provided $\lambda > L(\theta, \epsilon)$.

Proof of the lemma.[10] From the known integral representation[11]

$$P_n^{(\lambda)}(\cos\theta) = \int_0^{\pi} (\cos\theta + i \sin\theta \cos\phi)^n \sin^{2\lambda-1}\phi \, d\phi,$$

we have

$$\Delta_n^{\lambda} = p_n^{\lambda}(\cos\theta) - \cos^n\theta = \int_0^{\pi} F_n(\theta, \phi) \sin^{2\lambda-1}\phi \, d\phi \Big/ \int_0^{\pi} \sin^{2\lambda-1}\phi \, d\phi, \tag{4.7}$$

where

$$F_n(\theta, \phi) = (\cos\theta + i \sin\theta \cos\phi)^n - \cos^n\theta. \tag{4.8}$$

Evidently

$$|\, F_n(\theta, \phi) \,| \leqq 2. \tag{4.9}$$

Now we choose a $\delta = \delta(\theta)$ such that

$$0 < \delta < \tfrac{1}{2}\pi, \qquad \cos^2\theta + \sin^2\theta \sin^2\delta < 1. \tag{4.10}$$

From (4.7), (4.8) and (4.9), we have

$$\begin{aligned} |\, \Delta_n^{\lambda} \,| &\leqq 4 \int_0^{\frac{1}{2}\pi-\delta} \sin^{2\lambda-1}\phi \, d\phi \Big/ \int_0^{\pi} \sin^{2\lambda-1}\phi \, d\phi \\ &\qquad + \int_{\frac{1}{2}\pi-\delta}^{\frac{1}{2}\pi+\delta} |\, F_n(\theta, \phi) \,| \sin^{2\lambda-1}\phi \, d\phi \Big/ \int_0^{\pi} \sin^{2\lambda-1}\phi \, d\phi \\ &\leqq 4 \int_0^{\frac{1}{2}\pi-\delta} \sin^{2\lambda-1}\phi \, d\phi \Big/ \int_0^{\pi} \sin^{2\lambda-1}\phi \, d\phi \\ &\qquad + (\cos^2\theta + \sin^2\theta \sin^2\delta)^{\frac{1}{2}n} + |\cos\theta|^n. \end{aligned}$$

Now let ϵ be given. Let $n_0 = n_0(\theta, \epsilon)$ be such that $n > n_0$ implies

$$(\cos^2\theta + \sin^2\theta \sin^2\delta)^{\frac{1}{2}n} + |\cos\theta|^n < \tfrac{1}{2}\epsilon.$$

[10] This elegant proof of Lemma 1 is due to Professor Szegö. My original proof was more complicated and based on a new estimate which I shall state here since it might possibly be used elsewhere. For $0 < \theta < \frac{1}{2}\pi$, $\lambda \geqq 5$, and $n = 1, 2, 3, \cdots$ the following inequality holds

$$|\, P_n^{(\lambda)}(\cos\theta) \,| \leqq P_n^{(\lambda)}(1) \left\{ \left(\frac{1 + \cos^2\theta}{2} \right)^{\frac{1}{2}n} + \frac{4}{n} \frac{1 + \cos^2\theta}{\sin(2\theta)} \right\}.$$

[11] See [4], formulas (9) and (11), pp. 193–194.

The existence of such an n_0 is assured by (4.10). Furthermore, suppose $\lambda_0 = \lambda_0(\theta, \epsilon)$ is such that $\lambda > \lambda_0$ implies

$$4\int_0^{\frac{1}{2}\pi-\vartheta} \sin^{2\lambda-1}\phi\, d\phi \Big/ \int_0^{\pi} \sin^{2\lambda-1}\phi\, d\phi < \tfrac{1}{2}\epsilon.$$

Hence $|\Delta_n^\lambda| < \epsilon$, provided $\lambda > \lambda_0(\theta, \epsilon)$, $n > n_0(\theta, \epsilon)$. On the other hand, from (4.4) we get that $|\Delta_n^\lambda| < \epsilon$ for $n = 0, 1, \cdots, n_0(\theta, \epsilon)$, provided $\lambda > \lambda_1(\theta, \epsilon)$. Now the lemma is proved if we take as $L(\theta, \epsilon)$ the larger of the two numbers λ_0 and λ_1.

Proof of Theorem 2. Our starting point is the expansion (4.3). Since $p_n^\lambda(1) = 1$ and hence

$$f(1) = \sum_{n=0}^{\infty} b_n(\lambda),$$

we see that the coefficients $b_n(\lambda)$ are uniformly bounded for all n and the range of values of λ. By Cantor's diagonal process we may find a subsequence $\lambda_\nu \to \infty$ such that

$$\lim_{\nu\to\infty} b_n(\lambda_\nu) = a_n \geqq 0, \qquad (n = 0, 1, 2, \cdots). \tag{4.11}$$

Let θ have a fixed value between 0 and π. Let us write the relation (4.3) in the form

$$f(\cos\theta) = \sum_0^\infty b_n(\lambda_\nu)\cos^n\theta + \sum_0^\infty b_n(\lambda_\nu)[p_n^{\lambda_\nu}(\cos\theta) - \cos^n\theta].$$

Since by the lemma we have

$$\left|\sum_0^\infty b_n(\lambda_\nu)[p_n^{\lambda_\nu}(\cos\theta) - \cos^n\theta]\right| < \epsilon\sum b_n(\lambda_\nu) = \epsilon f(1),$$

provided λ_ν is sufficiently large, we may write

$$f(\cos\theta) = \sum_0^\infty b_n(\lambda_\nu)\cos^n\theta + \sigma, \tag{4.12}$$

where $|\sigma| < \epsilon f(1)$ for sufficiently large λ_ν. However, the series

$$\sum_0^\infty b_n(\lambda_\nu)\cos^n\theta \tag{4.13}$$

converges uniformly with respect to the variable λ_ν because it is majorized by the convergent series with constant terms

$$\sum_0^\infty f(1)\,|\cos\theta|^n.$$

But now the limiting relations (4.11) imply that the series (4.13) will tend to $\sum a_n\cos^n\theta$ as $\lambda_\nu \to \infty$. Thus (4.12) may be written as

$$f(\cos\theta) = \sum_0^\infty a_n\cos^n\theta + \sigma',$$

where $|\sigma'| \leqq \epsilon f(1)$. Now letting $\epsilon \to 0$, we conclude that $\sigma' = 0$ and hence that

$$f(\cos\theta) = \sum_0^\infty a_n \cos^n\theta, \qquad (a_n \geqq 0). \tag{4.14}$$

There remains to show that (4.14) is also valid for $\theta = 0$ and $\theta = \pi$. This last point, however, is readily settled. Indeed, (4.14) implies the convergence of the series $\sum a_n$ by letting $\theta \to 0$. Now the continuity of both sides of the relation (4.14) at both ends of the interval $0 \leqq \theta \leqq \pi$ implies its validity throughout this closed interval.

5. **On metric transforms of S_∞ which are imbeddable in Hilbert space.** Suppose $F(t)$, $(0 \leqq t \leqq \pi)$, is continuous, $F(0) = 0$, $F(t) \geqq 0$ if $0 < t \leqq \pi$. Let us remetrize the metric space S_∞ from the original distance function pq to the new distance function $F(pq)$. The new semi-metric space thus obtained is called the metric transform of S_∞ by the function $F(t)$ and denoted by the symbol $F(S_\infty)$. We shall prove the following theorem.[12]

THEOREM 3. *The metric transform $F(S_\infty)$ is isometrically imbeddable in H if and only if*

$$F^2(t) = \sum_{n=1}^\infty a_n(1 - \cos^n t), \qquad (a_n \geqq 0,\ 0 \leqq t \leqq \pi). \tag{5.1}$$

We need the following two lemmas.

LEMMA 2. *The metric transform $F(S_\infty)$ is imbeddable in H if and only if the function* $\exp\{-\lambda F^2(t)\}$ *is positive definite in S_∞ for all $\lambda > 0$.*

This lemma is known to be valid not only for S_∞ but for any separable metric space ([6], Theorem 1, p. 527).

LEMMA 3. *Let K denote the class of functions $\phi(x)$ of the form*

$$\phi(x) = \sum_{n=1}^\infty a_n(1 - x^n), \qquad (a_n \geqq 0,\ -1 \leqq x \leqq 1). \tag{5.2}$$

If $\{\phi_n(x)\}$ is a sequence of functions of this class and

$$\lim_{n\to\infty} \phi_n(x) = \phi_0(x), \qquad (-1 \leqq x \leqq 1), \tag{5.3}$$

where $\phi_0(x)$ is continuous in $-1 \leqq x \leqq 1$, then also $\phi_0(x)$ belongs to the class K.

[12] A similar theorem concerning S_m is as follows. *$F(S_m)$ is imbeddable in H if and only if*

$$F^2(t) = \sum_{n=1}^\infty a_n(1 - p_n^\lambda(\cos t)), \qquad (\lambda = \tfrac{1}{2}(m-1),\ 0 \leqq t \leqq \pi).$$

However, this theorem is implied by a general result of Bochner ([2], Theorem 3) concerning a certain type of compact spaces, of which the simplest example is the finite dimensional sphere S_m. Our method of proving Theorem 3 was used before in the case of Euclidean and Hilbert spaces ([7], §5, [3], Part III).

Proof of Lemma 3. We prove first the following statement. Suppose

$$\psi(x) = \sum_0^\infty c_\nu x^\nu, \qquad (c_\nu \geqq 0, -1 < x < 1),$$

such that

$$\int_0^{1-0} \psi(x)\, dx \text{ exists.}$$

Then the class K is identical with the class K_1 of functions of the form

$$\phi(x) = \int_x^{1-0} \psi(x)\, dx, \qquad (-1 \leqq x < 1), \qquad \phi(1) = 0. \tag{5.4}$$

Proof that $K \subset K_1$. From $\phi(x) = \sum_1^\infty a_n(1 - x^n)$ we get $-\phi'(x) = \sum_1^\infty n a_n x^{n-1} = \psi(x)$ for $-1 < x < 1$; hence

$$\phi(x) - \phi(1 - \epsilon) = \int_x^{1-\epsilon} \psi(x)\, dx.$$

Letting $\epsilon \to 0$ we get (5.4); hence $\phi(x) \in K_1$.

Proof that $K_1 \subset K$. Suppose $\phi(x) \in K_1$, and hence is of the form (5.4). We have

$$\phi(x) - \phi(1 - \epsilon) = \int_x^{1-\epsilon} \left(\sum_0^\infty c_\nu x^\nu \right) dx = \sum_0^\infty \frac{c_\nu}{\nu + 1} [(1 - \epsilon)^{\nu+1} - x^{\nu+1}].$$

Letting $\epsilon \to 0$ we get $\phi(x) = \sum_0^\infty (1 - x^{\nu+1}) c_\nu/(\nu + 1)$; hence $\phi(x) \in K$.

Returning to the assumptions of Lemma 3, let

$$\phi_n(x) = \sum_{\nu=1}^\infty a_{n\nu}(1 - x^\nu).$$

If $x = re^{i\theta}$ $(0 \leqq r < 1)$, we have

$$\phi_n(x) = \sum_\nu a_{n\nu}(1 - r^\nu e^{i\nu\theta}) = \sum_\nu a_{n\nu}(1 - r^\nu) + \sum_\nu a_{n\nu} r^\nu (1 - e^{i\nu\theta});$$

hence

$$|\phi_n(x)| \leqq \phi_n(r) + 2 \sum_{\nu=1}^\infty a_{n\nu} r^\nu. \tag{5.5}$$

But $r^\nu < A \cdot (1 - r^\nu)$ for all $\nu = 1, 2, \cdots$, for a certain $A = A(r)$. Thus (5.5) implies that

$$|\phi_n(x)| \leqq \phi_n(r) + 2A(r)\phi_n(r)$$

for $|x| = r$ and therefore also if $|x| \leqq r$. From (5.3) we now conclude that the $\phi_n(x)$ are uniformly bounded in every circle $|x| \leqq r$ $(r < 1)$. By the

Vitali-Porter convergence theorem, (5.3) implies that the $\phi_n(x)$ converge uniformly inside $|x| < 1$ and that $\phi_0(x)$ is therefore analytic and regular in $|x| < 1$. Since $\phi_n(x) \in K_1$, we know that $-\phi_n'(x)$ has all its derivatives non-negative at the origin. From $\phi_n^{(k)} \to \phi_0^{(k)}$ (uniformly inside $|x| < 1$) we conclude that also $-\phi_0'(x)$ has all derivatives non-negative at the origin; hence

$$\psi_0(x) = -\phi_0'(x) = \sum_0^\infty c_\nu x^\nu, \qquad (c_\nu \geqq 0, -1 < x < 1).$$

This also implies

$$\phi_0(x) - \phi_0(1-\epsilon) = \int^{1-\epsilon} \psi_0(x)\,dx.$$

Since $\phi_0(1-0) = \phi_0(1) = 0$ we get, by letting $\epsilon \to 0$,

$$\phi_0(x) = \int_x^{1-0} \psi_0(x)\,dx.$$

Hence, $\phi_0(x) \in K_1 = K$ and our proof is completed.

Proof of Theorem 3. In order to prove the direct part of the theorem, let $F(t)$ be defined by (5.1). Now, for $\lambda > 0$,

$$\exp\{-\lambda F^2(t)\} = \exp\left\{-\lambda \sum_0^\infty a_n(1-\cos^n t)\right\}$$

$$= \exp\left\{-\lambda \sum_0^\infty a_n\right\} \cdot \exp\left\{\lambda \sum_0^\infty a_n \cos^n t\right\}$$

evidently admits an expansion of the form

$$\exp\{-\lambda F^2(t)\} = \sum_0^\infty b_n(\lambda)\cos^n t, \qquad (0 \leqq t \leqq \pi,\quad b_n(\lambda) \geqq 0), \tag{5.6}$$

and is therefore p. d. in S_∞ by Theorem 2.

The converse part is proved as follows. Let $F(S_\infty)$ be imbeddable in H. By Lemma 2 and Theorem 2 the expansion (5.6) is valid for all $\lambda > 0$ and therefore

$$\Phi(t,\lambda) = \frac{1}{\lambda}(1 - \exp\{-\lambda F^2(t)\}) = \sum_0^\infty \lambda^{-1} b_n(\lambda)(1-\cos^n t).$$

However, $\Phi(t,\lambda) \to F^2(t)$ as $\lambda \to 0$ and $\Phi(t,\lambda)$ is a function of class K if regarded as a function of the variable $x = \cos t$. As the limit function $F^2(t)$ is continuous throughout $0 \leqq t \leqq \pi$, Lemma 3 implies that $F^2(t)$ is of the desired form (5.1).

Remarks. 1. If in (5.1) we set $a_1 = 2$, $a_2 = a_3 = \cdots = 0$, we get $F_1^2(t) = 2(1-\cos t) = 4\sin^2(\frac{1}{2}t)$; hence

$$F_1(t) = 2\sin\tfrac{1}{2}t.$$

By Theorem 3, $F_1(S_\infty)$ is imbeddable in H. This is a trivial result, for $F_1(S_\infty)$ is the sphere S_∞ with the spherical distance pq changed to the Euclidean chord-distance $2\sin(\frac{1}{2}pq)$, which is already imbedded in H.

By taking only the second term of the expansion (5.1), we get $F_2^2(t) = 1 - \cos^2 t = \sin^2 t$; hence

$$F_2(t) = \sin t.$$

It should be noticed that this remetrization identifies diametrically opposite points on the sphere S_∞. This identification occurs for a $F(t)$ defined by (5.1) if and only if $a_n = 0$ whenever n is odd.

2. Let $F(S_\infty)$ be imbeddable in H. We want to show that $F(S_\infty)$ may also be placed in some appropriate spherical shell of H. Indeed, let us adjoin to the space $F(S_\infty)$ a new point A such that, if $p \in F(S_\infty)$, its distance to A is $d(A, p) = r$. We want to determine the constant r so that the space $F(S_\infty) + A$ can be imbedded in H. This requires (see [3], p. 229) that

$$\sum_{i,k=1}^{N} \{r^2 + r^2 - F^2(p_i p_k)\} x_i x_k \geqq 0$$

for arbitrary points $p_1, \cdots, p_N$ of S_∞. By (5.1) this amounts to

$$\sum_{i,k=1}^{N} \left\{2r^2 - \sum_{n=1}^{\infty} a_n(1 - \cos^n p_i p_k)\right\} x_i x_k \geqq 0.$$

If we set

$$r = \left(\frac{1}{2}\sum_{1}^{\infty} a_n\right)^{\frac{1}{2}}, \tag{5.7}$$

our last inequality reduces to

$$\sum_{n=1}^{\infty} a_n \sum_{i,k=1}^{N} \cos^n (p_i p_k)\, x_i x_k \geqq 0,$$

which is correct since $\cos^n t$ is p. d. in S_∞. We state this result as a

COROLLARY. *The metric transform $F(S_\infty)$ of Theorem 3, if imbedded isometrically in H, will lie on a spherical shell of H of radius* (5.7).

BIBLIOGRAPHY

1. P. APPELL ET J. KAMPÉ DE FÉRIET, *Fonctions hypergéométriques et hypersphériques. Polynomes d'Hermite*, Paris, 1926.
2. S. BOCHNER, *Hilbert distances and positive definite functions*, Annals of Mathematics, vol. 42(1941), pp. 647–656.
3. JOHN VON NEUMANN AND I. J. SCHOENBERG, *Fourier integrals and metric geometry*, Transactions of the American Mathematical Society, vol. 50(1941), pp. 226–251.
4. N. NIELSEN, *Théorie des fonctions métasphériques*, Paris, 1911.
5. G. PÓLYA UND G. SZEGÖ, *Aufgaben und Lehrsätze aus der Analysis*, vol. 2, Berlin, 1925.
6. I. J. SCHOENBERG, *Metric spaces and positive definite functions*, Transactions of the American Mathematical Society, vol. 44(1938), pp. 522–536.
7. I. J. SCHOENBERG, *Metric spaces and completely monotone functions*, Annals of Mathematics, vol. 39(1938), pp. 811–841.
8. W. H. YOUNG, *A note on a class of symmetric functions and on a theorem required in the theory of integral equations*, Messenger of Mathematics, vol. 40(1910), pp. 37–43.

UNIVERSITY OF PENNSYLVANIA.

THE ANALYTIC PROLONGATION OF A MINIMAL SURFACE

By E. F. Beckenbach

1. **Introduction.** A classical theorem is the following.[1]

Theorem A. *If a minimal surface S cuts a plane Π orthogonally, then S is symmetric with respect to Π.*

We shall establish the following generalization of Theorem A.

Theorem B. *If a minimal surface S is bounded in part by an arc C of a curve that lies in a plane Π, and if S approaches Π orthogonally, then S can be continued analytically as a minimal surface across Π and the extended surface is symmetric with respect to Π.*

A similar generalization of the following classical result has been given by J. Douglas:[2] *if a minimal surface contains a straight line in its interior, then the straight line must be an axis of symmetry of the surface.*

The proofs of Theorem A which have been given depend essentially on the fact that the plane curve is an interior curve on S. To prove Theorem B, it would be sufficient, in virtue of Theorem A, to prove that S can be continued analytically across Π. Actually, though, we establish both the possibility of analytic continuation and the symmetry at the same time, so that in particular a proof of Theorem A is included in our proof of Theorem B.

It is to be noted (see (2.3)) that we do not assume the arc C to be analytic. We prescribe the behavior of only one of the coordinate components, $z(u, v)$, as the parameter point approaches an arbitrary point on a given segment ab of the boundary of the domain of definition D. Indeed we are assuming not even that the part of the boundary of S in question is an arc of curve but only that the boundary lies in a plane.

It follows as a *consequence* of Theorem B, however, that the boundary of S on Π must necessarily be an analytic arc.

It is further to be noted (see (2.4)) that we do not assume that the normal to S approaches a definite position as the parameter point approaches a fixed point on ab but only that the component of the normal perpendicular to Π approaches zero.

2. **Analytic formulation.** We shall denote by D the upper half of the u, v-plane, $v > 0$; by D' the lower half, $v < 0$; and by ab a fixed open segment

Received August 26, 1941.

[1] See, for instance, the author's paper, *Minimal surfaces in Euclidean n-space*, American Journal of Mathematics, vol. 55(1933), pp. 458–468.

[2] J. Douglas, *The analytic prolongation of a minimal surface over a rectilinear segment of its boundary*, Duke Mathematical Journal, vol. 5(1939), pp. 21–29; see also Proc. Nat. Acad. Sci. U. S. A., vol. 26(1940), pp. 215–221.

$a < u < b$, $v = 0$ of the u-axis. For a surface S with coordinate functions given by

$$x = x(u, v), \qquad y = y(u, v), \qquad z = z(u, v),$$

we shall denote the direction cosines of the normal to S by $X(u, v)$, $Y(u, v)$, $Z(u, v)$.

Suppose the functions $x(u, v)$, $y(u, v)$, $z(u, v)$ have the following properties:

(2.1) they are harmonic in D,

$$x_{uu} + x_{vv} = 0, \qquad y_{uu} + y_{vv} = 0, \qquad z_{uu} + z_{vv} = 0;$$

(2.2) they satisfy throughout D the relations

$$x_u^2 + y_u^2 + z_u^2 = x_v^2 + y_v^2 + z_v^2, \qquad x_u x_v + y_u y_v + z_u z_v = 0;$$

(2.3) for all $(u_0, 0)$ on ab, i.e., for $a < u_0 < b$, $v = 0$,

$$\lim_{(u,v)\to(u_0,+0)} z(u, v) = 0;$$

(2.4) for all $(u_0, 0)$ on ab, i.e., for $a < u_0 < b$, $v = 0$,

$$\lim_{(u,v)\to(u_0,+0)} Z(u, v) = 0.$$

We shall show that under these hypotheses, the functions $x(u, v)$, $y(u, v)$, $z(u, v)$ can be extended analytically across ab into D', in accordance with

(1) $\qquad x(u, -v) = x(u, v), \quad y(u, -v) = y(u, v), \quad z(u, -v) = -z(u, v).$

3. **Proof.** By (2.3) and the principle of symmetry for harmonic functions, $z(u, v)$ can be continued as a harmonic function across ab into D' in accordance with the third formula in (1).

Since $z(u, 0) = 0$ on ab, it follows that on ab we have

$$z_u = 0.$$

Further, z_v is continuous on ab. Hence, by (2.2), for $(u_0, 0)$ on ab,

(2) $$\lim_{(u,v)\to(u_0,+0)} (x_u^2 + y_u^2 - x_v^2 - y_v^2) = z_v^2(u_0, 0),$$

(3) $$\lim_{(u,v)\to(u_0,+0)} (x_u x_v + y_u y_v) = 0;$$

and by (2.4),

(4) $$\lim_{(u,v)\to(u_0,+0)} \frac{x_u y_v - x_v y_u}{x_u^2 + y_u^2 + z_u^2} = 0.$$

From (3), (4) and the identity

$$(x_u x_v + y_u y_v)^2 + (x_u y_v - x_v y_u)^2 \equiv (x_u^2 + y_u^2)(x_v^2 + y_v^2),$$

we have

(5) $\qquad (x_u^2 + y_u^2)(x_v^2 + y_v^2) = \eta^2(u, v) + \zeta^2(u, v)(x_u^2 + y_u^2 + z_u^2)(x_v^2 + y_v^2 + z_v^2),$

where

$$\lim_{(u,v)\to(u_0,+0)} \eta(u, v) = 0, \qquad \lim_{(u,v)\to(u_0,+0)} \zeta(u, v) = 0.$$

From (5) and the continuity of z_u and z_v on ab, it follows that we have

$$\lim_{(u,v)\to(u_0,+0)} (x_u^2 + y_u^2)(x_v^2 + y_v^2) = 0. \tag{6}$$

Now (2) and (6) imply

$$\lim_{(u,v)\to(u_0,+0)} (x_v^2 + y_v^2) = 0;$$

that is,

$$\lim_{(u,v)\to(u_0,+0)} x_v = \lim_{(u,v)\to(u_0,+0)} y_v = 0. \tag{7}$$

From (7) it follows that the harmonic functions $x_v(u, v)$ and $y_v(u, v)$ can be continued as harmonic functions across ab in accordance with

$$x_v(u, -v) = -x_v(u, v), \qquad y_v(u, -v) = -y_v(u, v). \tag{8}$$

Partial integration of the functions in (8) with respect to v yields the analytic continuation of $x(u, v)$, $y(u, v)$ in accordance with the first two formulas in (1).

4. **Remark (added in proof).** In the theorem of Douglas, it is shown that conditions (2.1), (2.2) and

$$\lim_{(u,v)\to(u_0,+0)} x(u, v) = 0, \qquad \lim_{(u,v)\to(u_0,+0)} y(u, v) = 0 \tag{9}$$

imply that $x(u, v)$, $y(u, v)$, $z(u, v)$ can be extended analytically across ab into D' in accordance with

$$x(u, -v) = -x(u, v), \quad y(u, -v) = -y(u, v), \quad z(u, -v) = z(u,v). \tag{10}$$

A simple proof follows. The first two equations in (10) follow from (9) and the principle of symmetry, so that on ab the functions x_u, y_u, x_v, y_v are continuous and $x_u = y_u = 0$. By (2.2), then,

$$\lim_{(u,v)\to(u_0,+0)} (z_u^2 - z_v^2) = x_v^2(u_0, 0) + y_v^2(u_0, 0) \geqq 0, \qquad \lim_{(u,v)\to(u_0,+0)} z_u z_v = 0;$$

consequently,

$$\lim_{(u,v)\to(u_0,+0)} z_v = 0,$$

so that z_v can be extended harmonically across ab in accordance with

$$z_v(u, -v) = -z_v(u, v).$$

Partial integration yields the third equation in (10).

THE UNIVERSITY OF MICHIGAN.

ADDITIVE FUNCTIONS AND ALMOST PERIODICITY

By Philip Hartman and Aurel Wintner

1. Let $f = f(n)$ be a function defined for $n = 1, 2, \cdots$. Its mean-value, $M(f)$, is usually defined by

$$\frac{f(1) + \cdots + f(n)}{n} \to M(f) \quad \text{as} \quad n \to \infty,$$

provided that this limit exists (this proviso should include that $M(f) \neq \pm\infty$). However, in certain connections, this definition of a mean-value turns out to be too vague. Correspondingly, the *Disquisitiones Arithmeticae* consider a more restrictive definition,[1] which today can be formulated as follows:

$$\frac{f(m+1) + \cdots + f(m+n)}{n} \to M(f) \quad \text{as} \quad n \to \infty$$

uniformly for all $m(= 1, 2, \cdots)$. Let $M(f)$ then be denoted also by $M^*(f)$.

Thus $M(f)$ exists and equals $M^*(f)$ whenever $M^*(f)$ exists. But $M^*(f)$ need not exist when $M(f)$ exists. The situation is illustrated by the following observations:

(i) If $f(n) = 1$ or $f(n) = n^{\frac{1}{2}}$ according as n is not or is a perfect square, then $M(f)$ exists. Hence $\limsup f(n) = \infty$ does not preclude the existence of $M(f)$ for an f. But this situation is changed if M is replaced by M^*, since $|f(n)| <$ const. is a necessary condition for the existence of $M^*(f)$. In fact, the existence of $M^*(f)$ means that, if $\epsilon > 0$ is arbitrary and if N_ϵ is suitably chosen, then

$$\left| \sum_{k=m}^{m+n-1} f(k) - nM^*(f) \right| < \epsilon n \quad \text{whenever} \quad n \geqq N_\epsilon,$$

where m is arbitrary. Hence, if $\epsilon = 1$ and $N = N_1$, then

$$\left| \sum_{k=m}^{m+N} f(k) - (N+1)M^*(f) \right| < N + 1, \qquad \left| \sum_{k=m+1}^{m+N} f(k) - NM^*(f) \right| < N$$

for every m. Consequently, for every m,

$$|f(m) - M^*(f)| < N + 1 + N.$$

Since $M^*(f)$ and N are independent of m, it follows that f is bounded.

(ii) Birkhoff's ergodic theorem states that, under his assumptions, the sequence of images of an arbitrary L-integrable function possesses an M-mean almost everywhere. But the theorem becomes false if M is replaced by M^*. This is clear from (i) since the L-integrable function can be chosen rather un-

Received September 30, 1941.

[1] C. F. Gauss, *Werke*, vol. 1, pp. 362-366.

bounded. Since Borel's law of large numbers is implied by Birkhoff's ergodic theorem, it will follow from (iii) below that there exist ergodic flows (and, as a matter of fact, mixtures) for which M cannot be replaced by M^* even in case of bounded functions.

(iii) If $f_t(n)$, where $0 < t < 1$, denotes the n-th digit in the (infinite) dyadic representation of t, then, according to Borel, $M(f_t)$ exists for almost all t (and equals $\frac{1}{2}$ for almost all t). On the other hand, $M^*(f_t)$ exists only when t is chosen in a certain set of measure 0. This is implied by the fact that the dyadic representation of almost every t contains arbitrarily long stretches of *both* digits, it being understood that every stretch contains only *one* of the digits.

(iv) Let t_0 denote the t-value defined by $f_t(n) = \frac{1}{2} + \frac{1}{2}\lambda(n)$, where λ is Liouville's factor (that is, $\lambda(n) = (-1)^l$, if n has exactly l prime factors, which need not be distinct). Then t_0 is "normal" in the sense of (iii). In other words, $M(f_{t_0})$ exists $(=\frac{1}{2})$ but $M^*(f_{t_0})$ does not exist. In fact, the prime number theorem is known to be equivalent to the assertion that $M(\lambda)$ exists $(= 0)$ and the non-existence of $M^*(\lambda)$ follows from other properties of the distribution of the primes. (If $\epsilon > 0$ is arbitrary, Riemann's hypothesis is known to be equivalent to $\lambda(1) + \cdots + \lambda(n) = O(n^{\frac{1}{2}+\epsilon})$ and can therefore be expressed by saying that t_0 is "normal" in the sense of the standard estimate $f_t(1) + \cdots + f_t(n) = \frac{1}{2}n + O(n^{\frac{1}{2}+\epsilon})$ of the successive sums of the dyadic digits of almost every t.)

(v) The requirement that the mean square deviation of an $f = f(n)$ from suitable finite trigonometric sums in n be arbitrarily small is equivalent to the almost periodicity of f in the sense of Besicovitch (B^2) or of Weyl (W^2) according as the mean is defined as M or as M^*. This is clear from the definitions in the first case and is easily verified in the second case.[2] That the replacement of M by M^* actually restricts the almost periodic class admitted, i.e., that not every f which is almost periodic (B^2) is almost periodic (W^2), is implied by the fundamental fact that the space (B^2) is, but the space (W^2) is not, complete with reference to the topology determined by the metric $M(|f_1 - f_2|^2)$, $M^*(|f_1 - f_2|^2)$ respectively.

(vi) An unbounded $f(n)$ can be almost periodic (B^2). It cannot be almost periodic (W^2). In fact, if an $f(n)$ is almost periodic (W^2), then $M^*(f)$ exists, and therefore $|f(n)| < \text{const.}$, by (i).

2. A function $f = f(n)$ is called additive if $f(n_1 n_2) = f(n_1) + f(n_2)$ whenever n_1 and n_2 are relatively prime (in particular $f(1) = 0$ for every additive f). Thus, if p denotes a prime number and k a positive integer, an additive f is uniquely determined by an arbitrary double sequence $\{\{c_p^{(k)}\}\}$ and by the assignments

$$f(p^k) = c_p^{(k)}$$

[2] H. Weyl, *Integralgleichungen und fastperiodische Funktionen*, Mathematische Annalen, vol. 97(1926), pp. 338–356 (end). It is understood that the functions $f(t)$, $-\infty < t < \infty$, considered there are now replaced by functions of the positive integer n. Cf. also B. Jessen and A. Wintner, *Distribution functions and the Riemann zeta function*, Transactions of the American Mathematical Society, vol. 38(1935), pp. 48–88 (§11–§12).

and

$$f(n) = \sum_{p^k \mid n} f(p^k), \qquad (p^{k+1} \nmid n). \tag{1}$$

It was recently shown[3] that an additive f is almost periodic (B^2) if and only if both series

$$\sum_p \frac{f(p)}{p}, \qquad \sum_p \sum_{j=1}^{\infty} \frac{|f(p^j)|^2}{p^j}$$

are convergent (the first of these series, which need not be absolutely convergent, being thought of as ordered according to increasing primes p). The main object of the present note is to show that *an additive f is almost periodic (W^2) if and only if it is bounded.* For an explicit criterion, cf. the end of §5 below.

Since the existence of $M^*(f)$ is necessary but not sufficient for the almost periodicity (W^2) of an arbitrary f, it is clear from (i) that the italicized theorem may be formulated as follows.

THEOREM. *If $f(n)$ is additive, the trivial necessary condition $|f(n)| < \text{const.}$ is sufficient not only for the existence of $M^*(f)$ but also for the almost periodicity (W^2) of $f(n)$; in particular, an additive $f(n)$ is almost periodic (W^2) whenever $M^*(f)$ exists.*

In view of (vi), only the sufficiency of this criterion needs a proof. Obviously, the theorem is purely arithmetical in nature; in fact, an arbitrary bounded $f(n)$ not only fails to be almost periodic (W^2) but does not even have a mean-value $M(f)$.

3. An arithmetical class of functions $g(n)$ considered by Toeplitz[4] (for $p = 2$) may be defined by choosing an arbitrary prime number p and a sequence of values $c_0, c_1, \cdots$, and then placing

$$g(p^k) = c_k$$

and

$$g(n) = g(p^k), \qquad \text{if } p^k \mid n \text{ but } p^{k+1} \nmid n, \qquad (k \geqq 0). \tag{2}$$

Let such a function g of the positive integer n be called a p-function (belonging to the fixed prime p).

It is known[5] that a p-function possesses a mean-value $M(g)$ if and only if

$$\sum_{k=0}^{\infty} \frac{g(p^k)}{p^k}$$

[3] P. Erdös and A. Wintner, *Additive functions and almost periodicity (B^2)*, American Journal of Mathematics, vol. 62(1940), pp. 635–645.

[4] O. Toeplitz, *Ein Beispiel zur Theorie der fastperiodischen Funktionen*, Mathematische Annalen, vol. 98(1928), pp. 281–295.

[5] E. R. van Kampen and A. Wintner, *On the almost periodic behavior of multiplicative number-theoretical functions*, American Journal of Mathematics, vol. 62(1940), pp. 613–626 (Theorem I).

is a convergent series. It will now be shown that $M^*(g)$ *exists if and only if the p-function g is bounded*:

$$|g(n)| < \text{const. for all } n. \tag{3}$$

According to (2), this is equivalent to

$$|g(p^k)| < \text{const. for all } k. \tag{4}$$

The necessity of (3) follows from (i), §1.

In the proof of the sufficiency of (3), let it be assumed, without loss of generality, that g is real-valued. For every pair n, m of positive integers, let a function $\phi_n^{(m)}(x)$, $-\infty < x < \infty$, be defined by placing

$$\phi_n^{(m)}(x) = \frac{1}{n} \sum_{\substack{g(i)<x \\ m<i\leq m+n}} 1 \tag{5}$$

so that $\phi_n^{(m)}(x)$ is the relative frequency ("probability") of the inequality $g(i) < x$ when the integer i is greater than m but not greater than $m + n$. Thus the function $\phi_n^{(m)}$ of x is monotone and such that

$$\phi_n^{(m)}(-\infty) = 0, \qquad \phi_n^{(m)}(\infty) = 1. \tag{6}$$

Furthermore, for every non-negative integer h,

$$\int_{-\infty}^{\infty} x^h \, d\phi_n^{(m)}(x) = \frac{1}{n} \sum_{i=m+1}^{m+n} (g(i))^h. \tag{7}$$

If $k_1 = k_1(x)$, $k_2 = k_2(x)$, $\cdots$, where $k_1 < k_2 < \cdots$, denotes the sequence (finite or infinite) of those integers $k = k(x)$ which satisfy the inequality $g(p^k) < x$, it is readily verified from (2) and (5) that

$$\phi_n^{(m)}(x) = \frac{1}{n} \sum_j \left\{ \left[\frac{m+n}{p^{k_j(x)}} \right] - \left[\frac{m}{p^{k_j(x)}} \right] - \left[\frac{m+n}{p^{k_j(x)+1}} \right] + \left[\frac{m}{p^{k_j(x)+1}} \right] \right\},$$

where $[y]$ denotes the integral part of y and the summation index runs through all the subscripts of $k_j = k_j(x)$. Since this representation of $\phi_n^{(m)}(x)$ obviously implies that, for every fixed N,

$$\left| \phi_n^{(m)}(x) - \left(1 - \frac{1}{p}\right) \sum_{j<N} \frac{1}{p^{k_j(x)}} \right| < \frac{4N}{n} + \frac{1}{p^{k_N(x)}} + \frac{2}{n},$$

it follows that the limit relation,

$$\phi_n^{(m)}(x) \to \left(1 - \frac{1}{p}\right) \sum_j \frac{1}{p^{k_j(x)}} \quad \text{as} \quad n \to \infty,$$

holds uniformly for all $m(= 1, 2, \cdots)$. But the definition of the $k_j(x)$ shows that the expression on the right of this limit relation is identical with the monotone function $\phi(x)$, $-\infty < x < \infty$, defined by

$$\phi(x) = \left(1 - \frac{1}{p}\right) \sum_{g(p^k)<x} \frac{1}{p^k}, \qquad (\phi(-\infty) = 0, \ \phi(\infty) = 1), \tag{8}$$

where the summation index, k, runs through those non-negative integers for which $g(p^k) < x$. Accordingly,

$$\phi_n^{(m)}(x) \to \phi(x) \quad \text{as} \quad n \to \infty \tag{9}$$

holds uniformly for all m. It is understood that x is arbitrarily fixed in such a way as to be distinct from the values contained in the sequence of the discontinuity points of the monotone function $\phi(x)$. In other words, the arrow in (8) is meant in the sense of the theory of monotone functions.[6]

All of this was independent of the assumption (3). Suppose now that (3) is satisfied. Then, by (5),

$$\phi_n^{(m)}(x) \equiv \begin{cases} 0 & \text{if } -\infty < x < -\text{const.}, \\ 1 & \text{if const.} < x < \infty; \end{cases} \tag{10}$$

so that the actual domain of integration on the left of (7) is uniformly bounded in n and m together. It follows therefore from Helly's theorem on term-by-term integration, that, since (9) holds uniformly in m, the limit relation,

$$\int_{-\infty}^{\infty} x^h \, d\phi_n^{(m)}(x) \to \int_{-\infty}^{\infty} x^h \, d\phi(x) \quad \text{as} \quad n \to \infty, \tag{11}$$

holds uniformly in m for every fixed h. Hence it is seen from (7) that $M^*(g^h)$ exists for every h. Since the existence of $M^*(g)$ follows by choosing $h = 1$, the proof is complete.

On choosing $h = 1$ in (7) and (11) and substituting (8) in the integral on the right of (11), one obtains for $M^*(g)$ the explicit representation

$$M^*(g) = \left(1 - \frac{1}{p}\right) \sum_{k=0}^{\infty} \frac{g(p^k)}{p^k}, \tag{12}$$

which will be needed in §4.

4. It is known that, if $g = g(n)$ is a p-function, it is almost periodic in the sense of Bohr if and only if

$$\lim_{k \to \infty} g(p^k)$$

exists.[7] It is also known that g is almost periodic (B^2) if and only if the series

$$\sum_{k=0}^{\infty} \frac{|\, g(p^k)\,|^2}{p^k}$$

is convergent.[8]

[6] Cf. A. Wintner, *On the asymptotic repartition of the values of real almost periodic functions*, American Journal of Mathematics, vol. 54(1932), pp. 339–345.

[7] This criterion is contained in the considerations of Toeplitz, loc. cit., footnote 4.

[8] E. R. van Kampen and A. Wintner, Theorem II, see footnote 5. Toeplitz (see footnote 4) has stated without proof that this criterion is necessary and sufficient for almost periodicity (B^2). If this statement were correct, it would follow that the two almost periodic classes (W^2), (B^2) are identical in the present case. But the criterion proved above implies that such is not the case. Toeplitz is not responsible for the erroneous statement; cf., in fact, R. Schmidt, *Die trigonometrische Approximation für eine Klasse von verallgemeinerten fastperiodischen Funktionen*, Mathematische Annalen, vol. 100(1928), pp. 334–356, p. 335, footnote 5.

It will now be shown that *g is almost periodic (W^2) whenever $M^*(g)$ exists; so that* (4) *is sufficient (and necessary) for the almost periodicity (W^2) of a p-function g.* Since (4) is equivalent to (3), the statement is, in the main, a particular case of the theorem italicized in §2 (actually, the statement is not implied by the wording of the theorem of §2, since not every p-function is additive); cf. §5 below.

In order to prove the sufficiency of (3) for almost periodicity (W^2), define a sequence of functions $g^{(1)} = g^{(1)}(n)$, $g^{(2)} = g^{(2)}(n)$, $\cdots$ by placing

$$g^{(j)}(n) = g(n) \text{ for } 1 \leqq n \leqq p^j \quad \text{and} \quad g^{(j)}(n + p^j) = g^{(j)}(n) \text{ for every } n,$$

where $g = g(n)$ is the given p-function and j a positive integer. It is easily verified from the definition (2) of a p-function that the function $| g^{(j)} - g |^2$ of n is a p-function. Furthermore, $| g^{(j)} - g |^2$ is a bounded function of n, since $g(n)$ is supposed to be bounded. Hence, the criterion proved in §3 assures the existence of $M^*(| g - g^{(j)} |^2)$, if this criterion for g is applied to $| g - g^{(j)} |^2$. Moreover, if g in (12) is replaced by $| g - g^{(j)} |^2$,

$$M^*(| g - g^{(j)} |^2) = \left(1 - \frac{1}{p}\right) \sum_{k=0}^{\infty} \frac{| g(p^k) - g^{(j)}(p^k) |^2}{p^k}.$$

But it is clear from the definition of the functions $g^{(j)}$ of n that the value of the infinite series on the right tends to 0 as $j \to \infty$. Consequently,

$$M^*(| g - g^{(j)} |^2) \to 0 \quad \text{as} \quad j \to \infty.$$

Since every $g^{(j)}$ is a periodic function of n, it follows that g is almost periodic (W^2).

5. The theorem announced in §2 can now be proved as follows.

For every additive function $f = f(n)$ and for every prime p, define a function $g_p = g_p(n)$ by placing

$$g_p(n) = f(p^k) \quad \text{if} \quad p^k \mid n \text{ but } p^{k+1} \nmid n, \qquad (k \geqq 0). \tag{13}$$

It is clear from the definition (2) that g_p is a p-function. Furthermore, $g_p(1) = f(1)$, and so $g_p(1) = 0$, by (1). But it is clear from the definitions (1), (2) that a p-function is additive if and only if its value for $n = 1$ is 0. Hence, every g_p is additive.

Moreover, from (13) and (1),

$$f(n) = \sum_p g_p(n) \qquad \text{for every } n, \tag{14}$$

where p runs through all primes; it is understood that the infinite series (14) has only a finite number of non-vanishing terms for every fixed n.

Suppose now that the additive function $f(n)$ is bounded. Then, according to (13), each of the functions $g_p(n)$ of n is bounded and therefore, by §4, almost periodic (W^2). Hence, in order to prove that $f(n)$ is almost periodic, it is sufficient to ascertain that, in virtue of the assumption $| f(n) | < \text{const.}$, the infinite

series (14) is uniformly convergent for all n $(= 1, 2, \cdots)$. But it is clear from (1) and (14) that $|f(n)| < \text{const.}$ is equivalent to[9]

$$\sum_p \text{fin sup}\, |g_p(n)| < \infty, \tag{15}$$

where the fin sup refers to the least upper bound belonging to fixed p and variable n. Since (15) implies the uniform convergence of (14), the proof is complete.

It is clear from (14) that (15) is equivalent to

$$\sum_p \text{fin sup}_k |f(p^k)| < \infty. \tag{15'}$$

The merit of the formulation (15′) of the necessary and sufficient condition, $|f(n)| < \text{const.}$, for the almost periodicity (W^2) consists in the fact that the data $f(p^k)$ occurring in (15′) are independent of one another; $f(n)$ being defined by (1).

6. The result of §3 can now be transferred from a p-function to an additive function, as follows. *An additive function $f(n)$ has an M^*-mean if and only if is a bounded function.* For if $f(n)$ is additive and bounded, then it is almost periodic (W^2), and therefore $M^*(f)$ exists; while if $M^*(f)$ exists, $f(n)$ must be bounded, by (i), §1.

It follows that *the mere existence of $M^*(f)$ implies the almost periodicity (W^2) of $f(n)$, if $f(n)$ is additive.*

7. It is known[10] that an additive function $f(n)$ is almost periodic (B^λ), for a fixed $\lambda \geqq 1$, if and only if all four series

$$\sum_p \frac{f(p)}{p}, \qquad \sum_p \frac{\min(|f(p)|^2, 1)}{p}, \qquad \sum_p \sum_{k=2}^{\infty} \frac{|f(p^k)|^\lambda}{p^k}, \qquad \sum_{|f(p)|\geq 1} \frac{|f(p)|^\lambda}{p}$$

are convergent. This implies that, if an additive $f(n)$ is almost periodic (B^λ), it need not be almost periodic $(B^{\lambda+\epsilon})$. On the other hand, *an additive $f(n)$ is almost periodic (W^λ) either for every λ or for no λ, where $1 \leqq \lambda < \infty$.* In fact, the above proof of the criterion $|f(n)| < \text{const.}$ for the almost periodicity (W^λ) of $f(n)$ in the case $\lambda = 2$ obviously holds for every λ.

The formulation (15′) of $|f(n)| < \text{const.}$ now appears of particular interest since the content of the condition $|f(p^k)| < \text{const.}$ for the independent data

[9] This trivial equivalence is only a manifestation of the statistical independence of the terms of the series (14); cf. P. Erdös and A. Wintner, *Additive arithmetical functions and statistical independence*, American Journal of Mathematics, vol. 61(1939), pp. 713–721 (§12); cf. B. Jessen and A. Wintner, loc. cit., footnote 2 (Theorem 3).

[10] P. Hartman and A. Wintner, *On the almost periodicity of additive number-theoretical functions*, American Journal of Mathematics, vol. 62(1940), pp. 753–758.

$f(p^k)$ becomes evident. In fact, if $|f(p^k)| < \text{const.}$ for all p and k, the four series quoted above are convergent for a fixed λ if and only if the three series

$$\sum_p \frac{f(p)}{p}, \qquad \sum_p \frac{\min(|f(p)|^2, 1)}{p}, \qquad \sum_{|f(p)|\geq 1} \frac{|f(p)|}{p}$$

are convergent. Since the latter do not involve λ, it follows that f is almost periodic (B^λ) either for every λ or for no λ, if $|f(p^k)| < \text{const.}$ Since all these conditions together do not assure that $|f(n)| < \text{const.}$, it also is seen that an additive f can be almost periodic (B^λ) for every λ without being almost periodic (W).

U. S. Army and The Johns Hopkins University.

A CORRECTION TO A PREVIOUS PAPER

By Charles B. Morrey, Jr.

1. **Introduction.** In a previous paper,[1] the author gave the following result (AC, Theorem 8.8):

A necessary and sufficient condition that the family $\{z(x)\}$ of functions of class $\mathfrak{P}_1$ on the bounded region G be compact with respect to weak convergence in $\mathfrak{P}_1$ on G is that the following two conditions hold:

(i) *$\bar{D}_1(z, G)$ is uniformly bounded;*

(ii) *there exists a non-negative convex function $\varphi(r_1, \cdots, r_n)$ with the property that*

$$\lim_{|r|\to\infty} |r|^{-1}\varphi(r_1, \cdots, r_n) = +\infty, \qquad |r|^2 = r_1^2 + \cdots + r_n^2,$$

and such that

$$\int_G \varphi[D_{x_1}z, \cdots, D_{x_n}z]\,dx$$

is uniformly bounded.

This result, in the generality stated, is false. However, it is possible to replace this result by other results which are sufficient for the applications which the author makes to the calculus of variations.

In this note, we shall use the notations and terminology of AC and shall assume that the reader is familiar with that paper. For purposes of clarity, however, we shall recall the definition of weak convergence in $\mathfrak{P}_1$ and a necessary and sufficient condition for compactness of a family with respect to weak convergence in $\mathfrak{P}_1$. One of the principal results of AC was that any of the spaces $\mathfrak{P}_\alpha$ (elements of which are classes of equivalent functions of class $\mathfrak{P}_\alpha$ on a bounded region G) are Banach spaces. Thus weak convergence in $\mathfrak{P}_\alpha$ is already defined in terms of that in a Banach space. A necessary and sufficient condition that a sequence $\{z_p(x)\}$ converge weakly on G to z in $\mathfrak{P}_\alpha$ is that $z_p \to z$ and $D_{x_i}z_p \to D_{x_i}z$ weakly in L_α on G $(i = 1, \cdots, n)$. The following necessary and sufficient condition for compactness with respect to weak convergence on a general bounded region G has been proved in AC.

Received October 20, 1941.

[1] *Functions of several variables and absolute continuity*, II, Duke Mathematical Journal, vol. 6(1940), pp. 187–215. Part I of this paper by J. W. Calkin appeared in the same issue of this journal, pp. 170–186. We shall hereafter refer to the two parts as one paper and shall denote it by the letters AC.

LEMMA (AC, Theorem 8.4). *If $\alpha > 1$, a necessary and sufficient condition that the family $\{z(x)\}$ be compact with respect to weak convergence in $\mathfrak{P}_\alpha$ on a bounded region G is that*

$$(1)\quad \bar{D}_\alpha(z, G) = \int_G |z|^\alpha\, dx + D_\alpha(z, G), \qquad D_\alpha(z, G) = \int_G \left[\sum_{i=1}^{n} (D_{x_i} z)^2\right]^{\frac{1}{2}\alpha} dx$$

be uniformly bounded. If $\alpha = 1$, we must add to the condition (1), the condition that the set functions

$$(2)\qquad \varphi(e) = \int_e |z|\, dx, \qquad \psi_i(e) = \int_e |D_{x_i} z|\, dx$$

be uniformly absolutely continuous on G.

The error made in proving the main result (AC, Theorem 8.8) was that no mention whatever was made of the set functions $\varphi(e)$. Actually, it was proved correctly that the conditions stated were necessary and sufficient for (1) to hold and for the set functions $\psi_i(e)$ to be uniformly AC. That (1) and the uniform absolute continuity of the $\psi_i(e)$ do not imply the uniform absolute continuity of the $\varphi(e)$ for an arbitrary region G is seen in the example in §2 below. We shall prove in Theorem 1 (§3) below that the result holds if G is of class K. We shall show in Theorem 2 (§4) that the result holds if weak convergence is replaced by a slightly weaker type of convergence. In Theorem 3 (§4) we show that this slightly weaker type of convergence coincides with ordinary weak convergence in the case the functions all vanish on G^* in the sense of AC, Definition 9.1, and hence in case the functions all coincide, in the sense of AC, Definition 9.3, on G^* with a particular function $z^*(x)$ of class $\mathfrak{P}_1$ on G.

2. **An example.** Let G be a bounded region in the (x, y)-plane consisting of the interior of the unit square $0 \leqq x \leqq 1, 0 \leqq y \leqq 1$, together with the interiors of a denumerable number of non-overlapping squares arranged in decreasing order of magnitude from left to right above the top of the unit square and each connected to it by a narrow rectangle with sides parallel to the axes. Let the unit square be denoted by Q_0, the other squares by $Q_1, Q_2, \cdots$, and the connecting rectangles by $R_1, R_2, \cdots$. Let A_p denote the area of Q_p, w_p the width of R_p, and h_p the height of R_p for each p; we assume that the A_p, w_p, and h_p all tend to zero and shall specify the rate at which we wish the w_p to tend to zero later.

On G, we define a sequence $\{z_p(x, y)\}$ of functions, each of class $\mathfrak{P}_1$ on G as follows:

$$z_p(x, y) = \begin{cases} 0, & (x, y) \in Q_0, \\ (y - 1)/h_p A_p, & (x, y) \in R_p, \\ A_p^{-1}, & (x, y) \in Q_p, \\ 0, & (x, y) \in Q_q + R_q, \qquad q \neq p. \end{cases}$$

Then $z_p(x, y) \geqq 0$, $D_x z_p = 0$, and $D_y z_p \geqq 0$ everywhere on G (except at the top and bottom of R_p where $D_y z_p$ does not exist) and we have

$$\iint_G z_p(x, y)\, dx\, dy = 1 + \frac{(h_p w_p)}{2A_p}, \qquad \iint_G D_y z_p\, dx\, dy = w_p A_p^{-1}.$$

If we merely choose w_p so that $w_p \cdot A_p^{-1} \to 0$, we see that $z_p(x, y)$ tends strongly in $\mathfrak{P}_1$ to the function $z(x, y) \equiv 0$ on any region Γ with $\bar{\Gamma} \subset G$. Moreover,

$$\lim_{p\to\infty} D_1(z_p, G) = \lim_{p\to\infty} \iint_G D_y z_p\, dx\, dy = 0$$

so that the derivatives $D_x z_p$ and $D_y z_p$ tend strongly to zero in L_1 on G.

Now, choose $\epsilon > 0$ and choose P so large that $D_1(z_p, G) < \epsilon$ for $p > P$. Then choose $\delta > 0$ so small that $|\psi_{1,p}(e)|$ and $|\psi_{2,p}(e)| < \epsilon$ for any e with $m(e) < \delta$, $p = 1, \cdots, P$. From the condition on P and from the fact that $D_x z_p$, $D_y z_p \geqq 0$, it follows that if $m(e) < \delta$ then $|\psi_{1,z}(e)|$, $|\psi_{2,p}(e)| < \epsilon$ for all p. Thus the set functions $\psi_{1,p}(e)$ and $\psi_{2,p}(e)$ are uniformly absolutely continuous. Furthermore $\bar{D}_1(z_p, G) = 1 + w_p \cdot A_p^{-1} \cdot (1 + h_p/2)$ which is uniformly bounded. However, since

$$\iint_{Q_p} |z_p(x, y)|\, dx\, dy = 1$$

for every p and $\lim_{p\to\infty} A_p = 0$, it follows that the set functions $\varphi_p(e)$ are not uniformly AC.

3. **The result for regions of class K.** In this section, we prove that our main result holds in case G is of class K. To do this, we first prove the following more general theorem:

THEOREM 1. *Let $\{z(x)\}$ be a family of functions defined and of class $\mathfrak{P}_1$ on a bounded region G of class K and suppose that $\bar{D}_1(z, G)$ is uniformly bounded. Then the set functions $\varphi(e)$ are uniformly* AC *on G.*

Proof. Let (Γ, γ, N, T) be a canonical covering of $\bar{G}$ (see AC, Definition 7.3), let

$$z(x) = \sum_{i=1}^{N} z_i(x)$$

be the corresponding canonical resolution of $z(x)$ (see AC, Definition 7.4), and let $w_i(y)$ be the transform of $z_i(x)$ under T_i, $i = 1, \cdots, N$. In case $w_i(y)$ is defined only on R_2 ($|y_j| < 1, j = 1, \cdots, n-1, -1 < y_n \leqq 0$) extend it to the whole of R_1 ($|y_j| < 1$) by the equation $w_i(y_n', y_n) = w_i(y_n', -y_n)$. Then each w_i is of class $\mathfrak{P}_1$ on R_1 and vanishes on R_1^* and we evidently have that

$\bar{D}_1(w_i, R_1)$ is uniformly bounded for each i (see AC, Theorem 6.1 and Lemma 7.4). From AC, Theorem 9.3, it follows that

$$\int_E |w_i(y)|\,dy \leqq \gamma_n^{-1/n}\cdot[m(E)]^{1/n}\cdot D_1(w_i, R_1), \qquad (\gamma_n r^n = m[C(P, r)])$$

for all measurable sets E in R_1. Since each T_i is regular and class K from R_1 or R_2 to Γ_i and since each $z_i(x)$ is zero on and near $\overline{G - \Gamma_i}$, it follows that

$$\int_e |z_i(x)|\,dx \leqq K\cdot[m(e)]^{1/n}\cdot D_1(z_i, G),$$

where K is a constant depending only on G. The result follows by addition.

Our first result is an immediate consequence of this theorem.

COROLLARY. *The principal theorem* (AC, Theorem 8.8) *holds if G is of class K.*

4. **The result, using a new type of convergence.** It is evident from the result of the preceding section that the conditions in our main theorem imply that the family $\{z(x)\}$ is compact with respect to weak convergence on each region Γ with $\bar{\Gamma} \subset G$. This suggests the following definition of "pseudo-weak convergence" in $\mathfrak{P}_1$ on G.

DEFINITION. We say that a sequence $\{z_p(x)\}$ *converges pseudo-weakly* to $z(x)$ in $\mathfrak{P}_1$ on G if $z_p(x)$ and $z(x)$ are of class $\mathfrak{P}_1$ on G and

(i) $\bar{D}_1(z_p, G)$ is uniformly bounded,

(ii) the set functions $\psi_{i,p}(e)$ are uniformly AC on G, and

(iii) $z_p(x)$ tends weakly in $\mathfrak{P}_1$ to $z(x)$ on each bounded region Γ with $\bar{\Gamma} \subset G$.

With this definition in mind we see immediately that the following theorem holds.

THEOREM 2. *Our principal result holds if the words "weak convergence" are replaced by "pseudo-weak convergence".*

Proof. For, given any sequence $\{z_p(x)\}$ of our family, we may choose a sequence of regions $\{G_k\}$, each of class K such that $\bar{G}_k \subset G_{k+1}$ for each k and any closed set F in G is interior to all the G_k for $k > k_F$. For each k, we may extract a subsequence $\{z_{q,k}\}$ of $\{z_{q,k-1}\}$ $(z_{q,0} = z_p)$ which converges weakly in $\mathfrak{P}_1$ on G_k to some function $z_k(x)$. Now, it is clear that $z_{k+1}(x) = z_k(x)$ (essentially) on G_k and that $z_k(x)$ is of class $\mathfrak{P}_1$ on G_k with

$$\bar{D}_1(z_k, G_k) \leqq \liminf_{q\to\infty} \bar{D}_1(z_{q,k}, G_k)$$

and so is bounded independently of k. Thus there is a function $z(x)$ of class $\mathfrak{P}_1$ on G which coincides with z_k on G_k for each k. If we now let $\{z_r(x)\}$ be the diagonal sequence, we see that $z_r(x)$ tends weakly in $\mathfrak{P}_1$ to $z(x)$ on each G_k and hence on each bounded region Γ with $\bar{\Gamma} \subset G$. This proves the theorem.

We now compare pseudo-weak convergence with weak convergence in case all the functions $z_p(x)$ vanish on G^*, G being bounded.

THEOREM 3. *Suppose each function $z_p(x)$ of a sequence $\{z_p(x)\}$ vanishes on G^* in the sense of* AC, *Definition* 9.1, *and suppose that $\{z_p(x)\}$ tends pseudo-weakly in $\mathfrak{P}_1$ on the bounded region G to a function $z(x)$. Then $\{z_p(x)\}$ converges weakly in $\mathfrak{P}_1$ on G to $z(x)$ and hence $z(x)$ vanishes on G^*.*

Proof. Since $\{z_p(x)\}$ tends pseudo-weakly in $\mathfrak{P}_1$ on G, we know that the set functions $\psi_{i,p}(e)$ are uniformly absolutely continuous over the whole of G and that $\{z_p(x)\}$ tends weakly in $\mathfrak{P}_1$ to $z(x)$ on each cell R with $\bar{R} \subset G$. Thus, from §1, it follows that $z_p(x)$ and $D_{x_i}z_p$ tend weakly in L_1 to z and $D_{x_i}z$ on each cell R with $\bar{R}$ in G. Also from AC, Theorem 9.3, it follows that

$$\int_e |\, z_p(x)\,|\, dx \leqq \gamma_n^{-1/n} \cdot [m(e)]^{1/n} \cdot D_1(z_p, G), \qquad e \subset G,$$

so that the set functions $\varphi_p(e)$ are also uniformly AC on G. But, from a well known theorem (AC, Lemma 8.1), this is sufficient for $z_p(x)$ and the $D_{x_i}z_p$ to tend weakly in L_1 on the whole of G to z and the $D_{x_i}z$, respectively. Thus z_p tends weakly in $\mathfrak{P}_1$ to z on the whole of G. That $z(x)$ also vanishes on G^* follows from AC, Theorem 9.2.

COROLLARY. *Theorem* 3 *still holds if all the $z_p(x)$ coincide in the sense of* AC, *Definition* 9.3, *on G^* with a particular function $z^*(x)$ of class $\mathfrak{P}_1$ on G; in this case $z(x)$ also coincides on G^* with $z^*(x)$.*

Proof. This follows immediately by considering the functions $z_p(x) - z^*(x)$ and $z(x) - z^*(x)$.

5. Further changes in AC, part II.

p. 192, line 4. For G read S.

p. 199, line 21. Insert "which are open in G" between "γ_N" and "such".

p. 203, second line above Lemma 8.1. For ([1], chapter IV) read ([1], chapter IV, pp. 64, 65 and chapter IX, pp. 135–136).

p. 210, footnote 4, first line. For = read $\leqq$.

p. 213, line 16. For ([1], chapter IX) read ([1], chapter IX, Theorem 2).

UNIVERSITY OF CALIFORNIA.

A GENERAL KUMMER THEORY FOR FUNCTION FIELDS

By Saunders Mac Lane and O. F. G. Schilling

INTRODUCTION

Kummer theory studies the generation of Abelian fields by radicals over a given base field F. This paper will develop a "relative" form of the theory, in which the base field F is a field of algebraic functions over a coefficient field. The modified theory then considers extensions of this F which are generated by radicals, or by arbitrary algebraic extensions of the coefficient field, or both. The normal extensions of this type, unlike the ordinary Kummer extensions, are in general non-Abelian.

This development was suggested by an attempt to generalize the principal ideal theorem of algebraic number theory. Schmidt, Hasse, and others [17], [8][1] have shown that the class field theory over algebraic number fields has a strict analogue for function fields of one variable over a finite coefficient field. One might then surmise that the principal ideal theorem of Hilbert [2], [7] has a similar analogue. We may phrase this conjecture as follows. Given the group of divisor classes of degree zero in F, does there exist an unramified Abelian extension K of F whose Galois group is isomorphic to the group of divisor classes and in which all these divisors become principal? We shall show that this is not the case.

First, we describe more explicitly the behavior of divisor classes which become principal in an extension, and show that the behavior of such principal divisors can be restated in an elementary fashion, free of arithmetic concepts.

Consider an algebraic function field F of one variable, over a field $\mathfrak{F}$ of constants (in the classical case, $\mathfrak{F}$ = the complex numbers). The field F has an abstract Riemann surface whose points P can be described as prime divisors; i.e., as homomorphic mappings of F on $\mathfrak{F}'$ plus ∞, where $\mathfrak{F}'$ is an algebraic extension of the field $\mathfrak{F}$ of constants. Each function x of F has a finite number of zeros and poles at various prime divisors P_i. With the proper multiplicities (positive for zeros, negative for poles) these may be listed as a formal product:

$$(x) = P_1^{e_1} P_2^{e_2} \cdots P_n^{e_n}.$$

This product is the *divisor* of x. Any such formal product $A = \prod P_i^{e_i}$ is called a divisor, though it need not be the divisor of any function x of the field. The sum $\sum e_i f_i$ is the *degree* of A, if $f_i = [\mathfrak{F}_i : \mathfrak{F}]$. The divisors of the form $A = (x)$ are called *principal* divisors; they always have degree zero (number of zeros = number of poles). If K is any finite algebraic extension of the given function field F, each divisor of F may be construed as a suitable divisor of K,

Received October 24, 1941; presented to the American Mathematical Society on May 2, 1941 under the title "The principal divisor theorem for function fields".

[1] Numbers in square brackets refer to the bibliography at the end of the paper.

so that in addition to divisors principal in F there may be divisors which first become principal in K.

The principal divisor problem is the following. Given a formal list of zeros and poles which do not correspond to any function x of F, construct a new algebraic function u which will have exactly these zeros and poles. Since the new function u generates an algebraic extension $K = F(u)$, the problem may be reformulated thus. Given in F a non-principal divisor A of degree zero, consider the structure of those algebraic extensions K in which A becomes principal.

Every divisor A which becomes principal in a finite extension K has a finite order in the sense that some integral power A^m of the divisor A is principal. To prove this it suffices to consider the case when the extension K is normal. In case F has a finite characteristic p, inseparable extensions must be included; hence, let K_0 be the field of all elements of K separable over F with a degree $n = [K_0:F]$. Suppose $A = (u)$ is a divisor principal in K. Then there is a power $q = p^e$ of the characteristic such that u^q is separable over F so that $A^q = (u^q)$ is principal in the separable subfield K_0. If N denotes the norm from K_0 to F, then $NA^q = A^{qn} = N(u)^q = (Nu^q)$, where Nu^q is an element of F. Therefore, A^{qn} is principal in F so that the given divisor A does have finite order. More specifically, this proves that every divisor principal in a normal extension has order dividing the degree (or, the "reduced degree" in the sense of Steinitz) of that extension.

The set $\mathfrak{D}$ of all divisors which become principal in a given extension K of F is a group under the natural formal multiplication of divisors. This group contains the subgroup (F) of all principal divisors, and the result just established asserts that every element of $\mathfrak{D}/(F)$ has finite order. The principal divisor problem then concerns the structure of those fields K for which the corresponding group $\mathfrak{D}$ contains a given group $\mathfrak{D}_0$ of divisors.

The arithmetic notion of a prime divisor (a point on the Riemann surface) is not essential to the statement of this problem. By reformulating the problem without using this concept, it is possible to treat also the case of function fields of any number of variables.

First, the principal divisors of F can be described directly. Two such divisors multiply by the rule $(x)(y) = (xy)$, so that the correspondence $x \to (x)$ maps the multiplicative group of non-zero functions x of F on the group of principal divisors. The functions with divisor $(x) = 1$ (i.e., with no zeros or poles) are simply the constants of $\mathfrak{F}$; hence the group of principal divisors may be described as the factor group $F^*/\mathfrak{F}^*$, where F^* (or $\mathfrak{F}^*$) denotes the group of all non-zero elements of the field F (or $\mathfrak{F}$). The group of principal divisors (u) of the extended field K may be described in similar fashion; it has as subgroup the group of principal divisors (x) of F, provided each principal divisor (x) of F is identified with the corresponding principal divisor in K.

Next, each non-principal divisor A of finite order may be replaced by a formal symbol d which will serve all the purposes discussed above, and which will be called a "fractional" divisor. Specifically, if (x) is any principal divisor of F

which is not a proper power of any other principal divisor, a fractional divisor $d = (x)^{1/m}$, for m a positive integer, is the generator d of any infinite cyclic group $\{d^i\}$ which contains the given principal divisor (x) as the power $d^m = (x)$. Two fractional divisors d and d' will be identified if there is an isomorphism between the corresponding groups which carries d into d' and leaves the principal divisor (x) fixed. In this sense there is one and only one fractional divisor $(x)^{1/m}$ for given x and m.

Consider now an (arithmetic) divisor A of order m in F with $A^m = (x)$. It can be identified with the fractional divisor $d = (x)^{1/m}$. The divisor A becomes principal in an extension K if and only if $A = (r)$, for some $r \in K$; that is, if and only if the corresponding fractional divisor d can be identified with the principal divisor $(r) = d$ of K. Every problem about divisors which become principal can thus be stated in terms of principal divisors and "fractional" divisors; the problems become more inclusive to the exact extent that some fractional divisors d do not correspond to arithmetic divisors A.

An element r which makes a divisor $d = A$ of F principal will be called a *radical* over F; that is, r is a radical if $(r)^m = (x)$ for some integer m and for some x in F. If K is any algebraic extension of F and $F(\{r\})$ is the subfield of K generated by all its radicals, then the divisors of F which become principal in K are exactly the divisors which become principal in $F(\{r\})$. In this sense, our problem reduces to the study of "radical" extensions of F which, like $F(\{r\})$, are generated by radicals. Observe, by the way, the simple properties of $F(\{r\})$ relative to K. It may be uniquely characterized as a maximal subfield of K which is a radical extension of F. Every subfield of K which is a radical extension of F is contained in $F(\{r\})$. If K is normal over F, so is $F(\{r\})$.

The decisive step of replacing the arithmetic divisors by the fractional divisors can be applied in other problems, such as the investigations of Deuring [6]. His paper studies the structure of Abelian extensions K of a function field F of one variable, in the special case when K has the same field of constants as does F (i.e., when each element of K algebraic over the original coefficient field $\mathfrak{F}$ necessarily lies in $\mathfrak{F}$) and when F contains all n-th roots of unity, where n is the degree of K/F and is prime to the characteristic of F. Since such a field K is a composite of cyclic fields over F, the elementary results of Galois theory in the solution of cyclic equations by radicals show that any such field K is a "radical" extension of F, in our sense. All of Deuring's main results on these Abelian extensions can be stated in terms of fractional divisors without using Deuring's elaborate study of the extensions of prime divisors. His principal theorems (Theorems 11, 12, and 13 in [6]) are essentially corollaries of our results below.

Our first chapter treats the algebraic structure of a radical extension. An extension of this type may be characterized by a relation between the degree of the extension and the order of the group of those divisors which become principal in the extension. From this it follows that any subextension of a radical extension can also be generated by radicals. The second chapter con-

cerns the Galois group of a radical extension. The major theorem states necessary and sufficient conditions for the realization of an abstractly given group as the Galois group of such an extension. To this end, the group is considered as an extension of a suitable normal subgroup by the corresponding factor group and the conditions for realization are stated in terms of the factor sets describing this group extension. The third chapter applies these results to more specific cases, such as the classical case of a finite coefficient field and the case of a divisor class of prime power order.

CHAPTER I

ALGEBRAIC PROPERTIES OF RADICAL EXTENSIONS

1. **Fractional divisors in general function fields.** The basis of our study is a fixed pair of fields $F \supset \mathfrak{F}$. We call $\mathfrak{F}$ the field of constants or *coefficient field*, and require that it be algebraically closed in F (i.e., that every element of F algebraic over $\mathfrak{F}$ lie in $\mathfrak{F}$). The transcendence degree of F over $\mathfrak{F}$ is arbitrary; for example, F might be a function field $F = \mathfrak{F}(x_1, \cdots, x_n, y_1, \cdots, y_m)$ of n variables over $\mathfrak{F}$, generated by n simultaneous indeterminates $x_1, \cdots, x_n$ and m elements $y_1, \cdots, y_m$ algebraic over $\mathfrak{F}(x_1, \cdots, x_n)$.

If x is any non-zero element (or "function") of F, the *principal divisor* (x) is the coset of x modulo the multiplicative subgroup $\mathfrak{F}^*$ of non-zero constants of $\mathfrak{F}$. Thus, $(x) = 1$ if and only if $x \in \mathfrak{F}^*$. The group of all principal divisors from F is the factor-group $F^*/\mathfrak{F}^*$, and will be denoted more briefly by (F). No element[2] of this group has finite order, for if $(x)^m = (x^m) = 1$, then $x^m = b$ is a constant of $\mathfrak{F}$, so that x is algebraic over $\mathfrak{F}$ and hence is in $\mathfrak{F}$ because of the hypothesis that $\mathfrak{F}$ is algebraically closed in F. Therefore, $(x)^m = 1$ implies $(x) = 1$.

For this reason, the correspondence $(x) \to (x)^m$ is an isomorphism between the group (F) and the subgroup $(F)^m$ of all m-th powers of elements of (F). The group $(F)^m$, isomorphic to (F), can thus be embedded in a larger Abelian group (F) in which every element has a unique m-th root. This embedding process can be applied to the original group (F) itself, letting m have the successive values $2!, 3!, \cdots, n!, \cdots$. The limit provides an Abelian group $D_\infty \supset (F)$ with the following properties:

(i) D_∞ has no elements of finite order.

(ii) Every element of D_∞ has finite order, modulo (F).

(iii) Every element of (F) has roots of all orders in D_∞; that is, for each (x) in (F) and each integer m, there exists d in D_∞ with $d^m = (x)$.

These three properties uniquely determine the group D_∞ up to isomorphism over (F). Furthermore, if D is any other Abelian group containing the base

[2] With the trivial exception of the identity element. This exception will not be mentioned in subsequent cases like this.

group (F) and having the first two properties, D can be isomorphically embedded[3] in D_∞ in one and only one way.

In the classical case where F is an algebraic function field of one variable over $\mathfrak{F}$, the group D_∞ consists of all those formal products $B = P^{e_1} \cdots P_n^{e_n}$ of prime divisors P_i with *rational* exponents e_i for which some power B^m is a principal divisor. Supported by this imperfect analogy, we thus appropriate classical terminology. The group $D_\infty \supset (F)$ will be the group of *fractional divisors* of $F/\mathfrak{F}$, or simply the group of *divisors*. The *order* of a divisor d of D_∞ is its order modulo (F), that is, the order of its coset $d(F)$. This coset is the *divisor class* of d. A *group of divisors* D is any subgroup of D_∞ which contains the group (F) of principal divisors. The exponent e of such a group D is the exponent of the Abelian group $D/(F)$; that is, it is the largest order of any divisor class in this group.

If the field K is an algebraic extension of the base field F, the corresponding script letter $\mathfrak{K}$ will denote the *field of constants* of K; that is, the field $\mathfrak{K}$ composed of all the elements of K which are algebraic over the original coefficient field $\mathfrak{F}$. Because of the transitivity of algebraic dependence, the field $\mathfrak{K}$ is algebraically closed in its function field K. Any function $x \neq 0$ of F has a principal divisor $(x) = (x; F)$ in F and a conceptually distinct principal divisor $(x; K)$ in K. But $(x; K) = 1$ if and only if $x = b$ is a constant of $\mathfrak{K}$. By virtue of the algebraic closure of $\mathfrak{F}$ in F, any constant b in both $\mathfrak{K}$ and F must lie in the original constant field $\mathfrak{F}$. Therefore, $(x; K) = 1$ if and only if $(x; F) = 1$; the correspondence $(x; F) \leftrightarrow (x; K)$ maps the principal divisors of F isomorphically on certain of the principal divisors of K. We can and will identify the divisor $(x) = (x; F)$ with the divisor $(x; K)$, so that (F) becomes a subgroup of (K).

Other principal divisors from K can sometimes be identified with fractional divisors from the group D_∞ of F. Specifically, an element r of K will be called a *radical* if some power $(r)^m$ of its principal divisor is a principal divisor of (F). The least such exponent m is the *order* of the radical r, and if $(r)^m = (x)$ for some function x of F, one must have $(r^m/x) = 1$, or r^m/x a constant b of the constant field $\mathfrak{K}$, so that

$$r^m = bx \qquad (r \text{ a radical}, x \in F, b \in \mathfrak{K}). \tag{1}$$

The group of principal divisors generated by (r) and (F) is an Abelian group without elements of finite order, such that every element has a finite order, modulo (F); thus it has an isomorphic replica within the group D_∞ of fractional divisors of F. Therefore, the principal divisor (r) may be interpreted as a fractional divisor $(r) = d$. In this sense, the fractional divisor d of F has become principal in the extended field K. Conversely, suppose that G is any group of principal divisors (z) of K which has an isomorphic embedding $(z) \rightarrow (z)'$ into the group D_∞, and let d be any divisor which becomes principal in this embedding, as $d = (z)'$. Then by the definition of D_∞, $d^m = (x)$ for some integer

[3] Here and subsequently isomorphisms of groups containing (F) are assumed to leave fixed all the elements of (F).

m, so that $(z)'^{m} = d^{m} = (x)$, and so $(z^{m}) = (x)$. The function z of the extended field is therefore a radical. If we identify each divisor (r) of a radical with the appropriate fractional divisor, the result may be stated as follows.

THEOREM 1. *A fractional divisor d of the field F can be interpreted as a principal divisor of the algebraic extension $K \supset F$ if and only if $d = (r)$, for some radical r of K/F.*

The radicals of K form a multiplicative group, the *group R of radicals*, which contains the subgroup $F^{*}\mathfrak{K}^{*}$ generated by non-zero elements of F and $\mathfrak{K}$. The group (R) of all principal divisors of radicals may be identified with a subgroup D of D_{∞}; this group consists of all fractional divisors of F which "become principal" in K, and we call it the *divisor group $D = D(K)$ associated* with K. The corresponding group $D/(F)$ of divisor classes is isomorphic to the quotient group $R/F^{*}\mathfrak{K}^{*}$. In sum, a field K with radical R and a divisor group D has

$$D = (R), \qquad D/(F) = (R)/(F) \cong R/F^{*}\mathfrak{K}^{*}. \tag{2}$$

Our object is the study of fields K with a given associated divisor group D. This study will reduce to that of "radical extensions"; we say that K is a *radical extension* of F if K can be generated from F by the adjunction to F of some or all of the radicals of K.

2. Pure coefficient extensions. It is essential to break an arbitrary extension $K \supset F$ into two stages $K \supset F(\mathfrak{K}) \supset F$, where the intermediate field $F(\mathfrak{K})$ is[4] generated by the adjunction to F of all the elements of the field $\mathfrak{K}$ of constants. A *coefficient extension* of F is any extension so generated. The requisite properties of these extensions will be stated now; most of them are well known, at least for function fields of one variable (see, for example, [6], §1).

Let $\mathfrak{W}$ be any *separable* algebraic extension of the original coefficient field $\mathfrak{F}$, and $W = F(\mathfrak{W})$ the coefficient extension generated by $\mathfrak{W}$. Because $\mathfrak{F}$ is algebraically closed in F, an equation for any element of $\mathfrak{W}$ irreducible over $\mathfrak{F}$ remains irreducible over F. If this fact is applied to a primitive element of a finite extension $\mathfrak{W}$ of $\mathfrak{F}$, the degree of W over F is found to be

$$[W:F] = [F(\mathfrak{W}):F] = [\mathfrak{W}:\mathfrak{F}]. \tag{3}$$

We now assert that $\mathfrak{W}$ is the coefficient field of $F(\mathfrak{W})$.

LEMMA 1. *$\mathfrak{W}$ is algebraically closed in $F(\mathfrak{W}) = W$.*

Proof. Consider first any element u of W algebraic and *separable* over $\mathfrak{W}$. There is then a subfield $\mathfrak{W}_{0} \subset \mathfrak{W}$ finite over $\mathfrak{F}$ with u in $F(\mathfrak{W}_{0})$. Two applications of (3) to $F(\mathfrak{W}_{0}) = F(\mathfrak{W}_{0}, u)$ give

$$[\mathfrak{W}_{0}(u):\mathfrak{F}] = [F(\mathfrak{W}_{0}, u):F] = [F(\mathfrak{W}_{0}):F] = [\mathfrak{W}_{0}:\mathfrak{F}];$$

[4] $F(\mathfrak{K})$ is simply the join $F \cup \mathfrak{K}$ of the two subfields F and $\mathfrak{K}$ of K. For such a join we use systematically the "adjunction" notation $F(\mathfrak{K})$.

hence, $\mathfrak{W}_0(u) \subset \mathfrak{W}_0$, so $u \in \mathfrak{W}$. Consider next an element $u \in W$ *inseparable* and algebraic over $\mathfrak{W}$. The irreducible equation for u over $\mathfrak{F}$ is then inseparable and as before remains irreducible and therefore inseparable over F. This contradicts the fact that $F(\mathfrak{W})/F$ is generated by a separable extension $\mathfrak{W}/\mathfrak{F}$.

Any automorphism S of W/F may be applied to the elements of $\mathfrak{W}$ contained in W and so induces an automorphism σ of $\mathfrak{W}/\mathfrak{F}$. If $\mathfrak{W}/\mathfrak{F}$ is finite and normal, consideration of the effects of this automorphism on a primitive element of $\mathfrak{W}/\mathfrak{F}$, which is also primitive for W/F, will prove the following result. By the usual directed systems of finite subfields this can be generalized to infinite normal extensions.

LEMMA 2. *If $\mathfrak{W}/\mathfrak{F}$ is normal, so is W/F. The correspondence $S \to \sigma$ carrying each automorphism of W/F into the induced automorphism σ of $\mathfrak{W}/\mathfrak{F}$ is an isomorphism of the Galois group of W/F to that of $\mathfrak{W}/\mathfrak{F}$. If $\mathfrak{W}/\mathfrak{F}$ is infinite, this isomorphism is continuous in both directions in the topology of the Galois groups.*

Here and subsequently we use the customary topology for the group of automorphisms of an infinite extension $W \supset F$: each neighborhood of 1 is determined by a subfield $M \subset W$ finite over F and consists of all automorphisms leaving M elementwise fixed.

Finally, we consider the effect of a simultaneous coefficient extension on a pair of fields $K \supset F$ with $\mathfrak{K} = \mathfrak{F}$.

LEMMA 3. *If K is a finite extension of F such that $\mathfrak{F}$ is algebraically closed in K while $\mathfrak{W}$ is a separable algebraic extension of $\mathfrak{F}$, then $K(\mathfrak{W})/F(\mathfrak{W})$ is finite of degree*

$$[K(\mathfrak{W}):F(\mathfrak{W})] = [K:F]. \tag{4}$$

Proof. If $\mathfrak{W}$ is finite over $\mathfrak{F}$, the equality (4) results at once from two applications of the equation (3) to the chains $K(\mathfrak{W}) \supset K \supset F$, $K(\mathfrak{W}) \supset F(\mathfrak{W}) \supset F$ joining $K(\mathfrak{W})$ to F. If $\mathfrak{W}$ is infinite over $\mathfrak{F}$, analyze K by the tower of fields, $F \subset K_0 \subset K_1 \subset \cdots \subset K_m = K$, where each K_e consists of all elements u of K whose p^e-th power is separable over F.

First, the separable extension K_0 is generated from F by a primitive element u. We already know that the polynomial equation for u over F remains irreducible over $F(\mathfrak{W}_0)$ for any *finite* subfield $\mathfrak{W}_0 \subset \mathfrak{W}$; hence, it is irreducible over the whole field $F(\mathfrak{W})$ and

$$[K_0(\mathfrak{W}):F(\mathfrak{W})] = [F(\mathfrak{W}, u):F(\mathfrak{W})] = [u:F] = [K_0:F]. \tag{5}$$

Second, the inseparable extension K_{e+1}/K_e of degree p^n and exponent p can be generated as $K_{e+1} = K_e(x_1, \cdots, x_n)$, where the elements $x_1^p, \cdots, x_n^p$ are p-independent in K_e. Since a separable extension $\mathfrak{W}/\mathfrak{F}$ preserves p-independence (see, e.g., [10]), these elements $x_1^p, \cdots, x_n^p$ remain p-independent in $K_e(\mathfrak{W})$ so that the extension $K_{e+1}(\mathfrak{W}) = K_e(\mathfrak{W}, x_1, \cdots, x_n)$ still has the degree p^n over $K_e(\mathfrak{W})$ or

$$[K_{e+1}(\mathfrak{W}):K_e(\mathfrak{W})] = [K_{e+1}:K_e] = p^n \qquad (n = n_e).$$

Combined with (5), this result for $e = 0, 1, 2, \cdots, m$ gives the desired conclusion (4).

In these lemmas the requirement that $\mathfrak{W}$ be separable over $\mathfrak{F}$ is not a captious restriction. For example, consider the field $\mathfrak{P}(x, y, z)$ generated by three independent indeterminates x, y, z over a perfect field $\mathfrak{P}$ of characteristic p, and let $\mathfrak{F} = \mathfrak{P}(x^p, y^p)$, $F = \mathfrak{P}(x^p, y^p, z, u) = \mathfrak{F}(z, u)$, where $u = yz + x$. To show that $\mathfrak{F}$ is actually algebraically closed in F, observe first that F is obtained from the pure transcendental extension $\mathfrak{F}(z)$ by the adjunction of a p-th root $u = (y^p z^p + x^p)^{1/p}$. If $u \in F$ is algebraic over $\mathfrak{F}$, $u^p \in \mathfrak{F}(z)$; hence, $u^p \in \mathfrak{F}$, by the nature of a transcendental extension, so that $u \in \mathfrak{P}(x, y)$. Thus u lies in the intersection of $\mathfrak{P}(x, y, z^p)$ and $\mathfrak{P}(x^p, y^p, z, u) = F$. In a study of certain non-modular lattices, it was shown [9] that this intersection is $\mathfrak{P}(x^p, y^p, z^p)$. But $u \in \mathfrak{P}(x^p, y^p, z^p)$ and u algebraic over $\mathfrak{P}(x^p, y^p)$ give u in $\mathfrak{P}(x^p, y^p)$ so that $\mathfrak{F}$ is indeed algebraically closed in F.

This field $\mathfrak{F}$ has a purely inseparable extension $\mathfrak{W} = \mathfrak{F}(y)$ of degree p ($y^p \in \mathfrak{F}$). The corresponding coefficient extension is $W = F(y) = \mathfrak{P}(x^p, y^p, z, yz + x, y)$, and contains an element $x = yz + x - y \cdot z$ algebraic over $\mathfrak{F}$ but not in $\mathfrak{W}$, so Lemma 1 fails. The equation (3) above will fail if one uses the extension $\mathfrak{W} = \mathfrak{F}(y, x)$ of degree p^2. One may also show that Lemma 3 would break down in a similar case.

At this point we may mention a somewhat more general type of radical extensions, in which the allowable coefficient extensions are restricted more sharply. Let any given field F be embedded in an algebraically complete[5] field A, and let $\mathfrak{C}$ be a specified subfield of A with the following properties:

(i) All roots of unity from A lie in $\mathfrak{C}$;

(ii) $\mathfrak{C}$ is normal and algebraic over $F \cap \mathfrak{C}$.

(For example, $\mathfrak{C}$ might simply be the subfield of A generated by all roots of unity.)

Coefficient extensions are now those extensions generated by subfields of $\mathfrak{C}$. Thus, we let $\mathfrak{F}$ denote the intersection $F \cap \mathfrak{C}$, while the field of constants of any algebraic extension K of F is to be found by embedding K in A, then taking the intersection $\mathfrak{K} = K \cap \mathfrak{C}$. Because of condition (ii), this intersection is independent of the way in which K may be embedded in A. With such coefficient extensions and with fields $\mathfrak{W}$ which are subfields of $\mathfrak{C}$, all the lemmas of this section are still true, provided the condition "$\mathfrak{W}$ algebraically closed in K" be replaced by the condition "$K \cap \mathfrak{C} = \mathfrak{W}$".

Relative to this field $\mathfrak{C}$, divisors and radical extensions can still be defined. By assumption (i) on $\mathfrak{C}$, the group $(F) = F^*/\mathfrak{F}^*$ of principal divisors of F still is an Abelian group without elements of finite order, contained in the larger group $K^*/\mathfrak{K}^*$ of principal divisors of K. Under this interpretation of the divisors, the whole subsequent theory of radical extensions will go through without change as a theory of extensions relative to $\mathfrak{C}$. In case $\mathfrak{C}$ is algebraically com-

[5] Algebraically complete = has no proper algebraic extension. We use this term in place of the older and less convenient one, "algebraically closed".

plete, this gives simply the previous case. If $\mathfrak{C}$ is not algebraically complete but if $\mathfrak{C} \cap F$ is algebraically closed in F, the theory relative to $\mathfrak{C}$ is sharper than the standard one in the sense that any radical extension of F relative to $\mathfrak{C}$ is always a radical extension in the standard sense, but not necessarily conversely.

3. **The inseparable cases.** Separable and inseparable radical extensions can be sharply distinguished.

THEOREM 2. *If the field F has the finite characteristic p, then every divisor of F principal in a separable extension of F has order relatively prime to p and every divisor principal in a purely inseparable extension of F has order a power of p. Conversely, if d is any divisor of F principal in some extension $K \supset F$, then, if d has order prime to p it is principal in some subextension of K separable over $F(\mathfrak{K})$, while if d has as order some power of p, it is principal in some subextension of K which is purely inseparable over $F(\mathfrak{K})$.*

Proof. Any divisor $d = (r)$ principal in K is principal in the extension of $F(\mathfrak{K})$ generated by a root r of an equation $r^m = xa$, irreducible over $F(\mathfrak{K})$. This equation is separable or purely inseparable according as the order m of d is prime to p or is a power of p. From this fact the various statements of the theorem follow.

This result means that the study of radical extensions in which the associated group D of divisors is a p-group (a group in which every element has order some power of p) is coextensive with the study of purely inseparable extensions. Specifically, an extension K with coefficient field $\mathfrak{K}$ is a radical extension of F with a divisor group which is a p-group if and only if K is a purely inseparable extension of $F(\mathfrak{K})$. For example, if F is a function field of one variable over a perfect coefficient field $\mathfrak{F}$, then F has one and only one purely inseparable extension of degree p^e, and in this extension all divisors of order p^e, and no others, become principal.

Henceforth, we omit these anomalous cases and treat only *separable* extensions K of F and divisor groups D in which the order of every element is *prime* to the characteristic p.

4. **Crossed characters.** The ideals which become principal in a normal extension of an algebraic number field F_0 give rise to certain functions defined on the Galois group of the extension. Specifically, if an ideal $\mathfrak{a}$ of F_0 is a principal ideal $\mathfrak{a} = (B)$ in some normal extension K_0 of F_0, then for each automorphism S of K_0 over F_0, the ideals (B) and (B^S) are equal so that each quotient B^S/B is a unit E_S of K_0. These units satisfy the functional equation $E_S(E_T)^S = E_{ST}$; hence they have been called "crossed characters". The properties of these functions have been studied by several authors;[6] various complications arise because the functions involve the explicit structure of the group of all units of the number field K_0. In the analogous situation for an algebraic function field, the units are

[6] See the references given for Theorem 10.3 in [11]. Note that we write $(B^T)^S = B^{ST}$.

replaced by the non-zero constants of the base field. Since these constants themselves constitute a field, the group structure should be more amenable to treatment. This gave the starting point of the present investigations.

To set up the corresponding functions for a general function field F over $\mathfrak{F}$, let G denote the group of automorphisms $u \to u^S$ of a separable normal extension K/F and observe that each automorphism $S \in G$ induces an automorphism $b \to b^S$ of the subfield $\mathcal{K}$ of all constants (although distinct automorphism S, T may induce the same automorphism of $\mathcal{K}$). A *crossed character* of G in $\mathcal{K}$ is a function which assigns to each S in G a non-zero constant c_S in $\mathcal{K}$ in such fashion that

$$c_S(c_T)^S = c_{ST} \qquad \text{(for all } S, T \text{ in } G\text{).} \tag{6}$$

The term-by-term product $c_S c_S'$ of two given crossed characters c_S and c_S' is again a crossed character, and the set of all crossed characters is a group $\{c_S\}$. If $b \neq 0$ is any fixed element of $\mathcal{K}$, the special function $c_S = b/b^S = b^{1-S}$ always satisfies (6); it is called a *unit character*.

Unless $K = F(\mathcal{K})$ is a pure coefficient extension, the Galois group of $\mathcal{K}/\mathfrak{F}$ is a *proper* homomorphic image of G. However, for the special case of a pure coefficient extension, Noether's principal genus theorem "in minimalen" [13], [18] may be phrased thus: *If $\mathcal{K}$ is a finite separable normal extension of $\mathfrak{F}$ with Galois group Γ, every crossed character of Γ in $\mathcal{K}$ is a unit character.* This result will be used repeatedly and is an essential tool for our investigations.

In general, in the Abelian group of all crossed characters, the cosets modulo the subgroup of unit characters are the so-called classes of *associate* crossed characters, and the corresponding factor group $\{c_S\}/\{b^{1-S}\}$ is thus the group of classes of crossed characters.

THEOREM 3. *If K, a finite separable normal extension of F, has a radical R and a Galois group G, then the group $(R)/(F)$ of divisor classes which become principal in K is isomorphic to the group of classes of crossed characters of G in the field $\mathcal{K}$ of constants.*

Proof. Each automorphism S of K/F may be regarded as an automorphism $(u) \to (u^S) = (u)^S$ of the group of principal divisors of K. By definition, any radical r of K has $(r)^m = (x)$ for a suitable integer m and a suitable function x of F. For any S, $(r^S)^m = (r^m)^S = (x)^S = (x) = (r)^m$, while $(r^S)^m = (r)^m$ implies that $(r^S) = (r)$, for there are no divisors of finite order. Therefore, the radicals r of K are included among those elements z of K for which $(z) = (z^S)$ for every automorphism S.

Conversely, consider any z with $(z) = (z^S)$, and let $y = \Pi_S z^S$ be the norm of z from K to F. If K has the degree m over F, the divisor (y) is then $\Pi_S(z^S) = (z)^m$. Therefore, z must be a radical of K. This fact we record as

COROLLARY 1. *In a finite separable normal extension of F the radicals are the elements z for which $(z^S) = (z)$ for every automorphism S.*

Return to the proof of the theorem and consider a radical z. Since the divisor of a function is 1 only when the function is a constant, each such z determines

a set of constants $c_S \in \mathfrak{K}$ with $z^S = c_S z$. From this one may show formally that the function c_S satisfies the equation (6); hence it is a crossed character. Conversely, any given crossed character c_S in $\mathfrak{K}$ may be regarded as a crossed character in the whole field K. Noether's principal genus theorem quoted above applies to K and asserts that c_S has the form $c_S = z/z^S$ for some element $z \neq 0$ of K. Then $(z) = (c_S z^S) = (z^S)$; hence, z is a radical. Therefore, the correspondence $z \to c_S$ carries the group R homomorphically onto the whole group of classes of crossed characters. One may verify that in this correspondence c_S is a unit character if and only if z is in the group $F^*\mathfrak{K}^*$, generated by constants and by functions of F. Therefore, the group of classes of crossed characters is isomorphic to the factor group $R/F^*\mathfrak{K}^*$. But this group, by (2), is isomorphic to the group $D/(F)$ of those divisor classes which become principal in K. This proves the theorem.

The result of the theorem may be applied in particular to a pure coefficient extension; the result is[7]

COROLLARY 2. *In an extension $W = F(\mathfrak{W})$ generated by a separable coefficient extension, no non-principal divisor of F becomes principal.*

Proof. If $\mathfrak{W}$ is not already normal, embed it in a separable field normal over $\mathfrak{F}$. If there is any change, more divisors would become principal in the enlarged field W; hence it suffices to prove the corollary in the case when $\mathfrak{W}$ is normal. By Lemma 2, the Galois group G of $\mathfrak{W}/\mathfrak{F}$ is then effectively identical with that of W/F so that the principal genus theorem asserts that every crossed character of G in $\mathfrak{W}$ is a unit character or that the group of classes of crossed characters has but one class. The theorem now shows that no divisors become principal.

We now turn to the other extreme case, where $\mathfrak{K} = \mathfrak{F}$.

COROLLARY 3. *If K is a finite separable normal extension with Galois group G over F and if the coefficient field $\mathfrak{F}$ is algebraically closed in K, then the number of divisor classes of F principal in K is finite and is a divisor of the number of characters of G which can be realized in the coefficient field $\mathfrak{F}$. If $\mathfrak{F}$ has a finite characteristic p, these numbers are both prime to p.*

Proof. In this case, the new coefficient field $\mathfrak{K}$ is just $\mathfrak{F}$ again and every automorphism of K acts as the identity on $\mathfrak{F}$. By the definition (6), every crossed character is just an ordinary character, that is, a function $c(S)$ with $c(S)c(T) = c(ST)$. The number of such functions is finite and their values are necessarily roots of unity. The number of principal divisor classes is bounded by the number of such functions with values in $\mathfrak{K}$; this clearly depends on the presence or absence of suitable roots of unity in $\mathfrak{K}$; an explicit formula could be derived. In any event, a field $\mathfrak{F}$ of characteristic p contains no p-th roots of unity other than 1; hence, it contains no characters of order p, as asserted above.

[7] A special case of this result was found by one of us (Schilling) in 1935; it was communicated to H. Reichardt, who generalized it to any function field of one variable.

To obtain similar limitations on the number of divisors which become principal in non-normal extension, we need the following observation on the effect of a pure coefficient extension upon the radical.

LEMMA 4. *If an extension K/F has a radical R, while a separable extension $\mathfrak{F}'$ of $\mathfrak{F}$ gives rise to an extension $K' = K(\mathfrak{F}')$ of $F' = F(\mathfrak{F}')$ with radical R', then $(R)/(F)$ is isomorphic to a subgroup of $(R')/(F')$.*

Proof. The groups (R), (F), and (F') are all subgroups of the group (R'). Any element in the intersection $(R) \cap (F')$ has the form (x'), where $x' \in F'$ and where $(x')^m$ is in (F), for some m. This means that x' is a radical of F' relative to F; since F' is a pure coefficient extension, Corollary 2 above shows that (x') is in (F). Hence $(R) \cap (F') = (F)$. Each coset of $(R)/(F)$ is then contained in one and only one coset of $[(R) \cup (F')]/(F')$. Since $(R) \cup (F') \subset (R')$, this coset correspondence is the required isomorphism of $(R)/(F)$ to a subgroup of $(R')/(F')$.

We can now state the basic theorem limiting the size of the radical.

THEOREM 4. *A finite separable extension $K \supset F$ with radical R has the order of its divisor group (R) bounded by*

$$[(R):(F)] \leqq [K:F(\mathcal{K})]. \tag{7}$$

In this theorem, the replacement of F by $F' = F(\mathcal{K})$, K by $K' = K(\mathcal{K}) = K$ and R by R' will not change the right-hand of (7) and will not decrease the left-hand of (7), because of Lemma 4. It thus suffices to prove the theorem in the case when $\mathcal{K} = \mathfrak{F}$.

If $\mathfrak{F} = \mathcal{K}$, the proof can be given by a two-step replacement of the given field K. Let $L = F(R)$ be the subfield generated by all radicals of K; it is a radical extension, with a radical Q which contains R. It is generated by roots r of various separable binomial equations $r^m = xa$. To enforce normality, let $\mathfrak{F}'$ be the (separable) extension of $\mathfrak{F}$ generated by all m-th roots of unity for every m occurring in such a binomial equation. The modified field $L' = F(\mathfrak{F}', R)$ will then be normal over $F' = F(\mathfrak{F}')$. Because $\mathcal{K} = \mathfrak{F}$, $\mathfrak{F}$ is algebraically closed in L; therefore, Lemma 1 proves $\mathfrak{F}'$ algebraically closed in $L' = L(\mathfrak{F}')$. Let Q' be the radical of L'/F'. One then has the following table:

	Given Extension		Radical Extension		Normal Extension
Top Field	K	$\supset$	$F(R) = L$	$\subset$	$L(\mathfrak{F}') = L'$
Base Field	F	$=$	F	$\subset$	$F(\mathfrak{F}') = F'$
Coef. Field	$\mathcal{K} = \mathfrak{F}$	$=$	$\mathfrak{F}$	$\subset$	$\mathfrak{F}'$
Radical	R	$\subset$	Q	$\subset$	Q'

FIGURE 1

The problem (see Figure 1) essentially reduces to the study of the normal extension L' of F', where the coefficient field of F' is algebraically closed in L'. By Corollary 3, the index $[(Q'):(F')]$ for its radical is bounded by the number

of characters of the Galois group which can be realized in $\mathfrak{F}'$. The number of characters is at most the order of the group, so this index is at most the degree of the field. Therefore, $[(Q'):(F')] \leqq [L':F']$. Tracing back the radicals through the construction used, we have, using Lemma 4,

$$[(R):(F)] \leqq [(Q):(F)] \leqq [(Q'):(F')],$$

while a similar description of the degrees, using Lemma 3, gives

$$[L':F'] = [L(\mathfrak{F}'):F(\mathfrak{F}')] = [L:F] = [F(R):F] \leqq [K:F].$$

These inequalities combine to prove $[(R):(F)] \leqq [K:F]$, where $F = F(\mathcal{K})$, as required by (7).

COROLLARY. *If K is a normal separable extension of F with constant field $\mathcal{K} = \mathfrak{F}$ and if a divisor D of order m becomes principal in K, then all m-th roots of unity lie in $\mathfrak{F}$.*

Proof. Let $D = (r)$ for r in K. Then r satisfies an equation $r^m = x$, for some x in F. By Theorem 4, the field $F(r)$ has degree at least m over F; hence this equation for r is irreducible over F. Since K is normal, the conjugates of r all lie in K; they are $\zeta^i r$ for ζ a primitive m-th root of unity. Hence $\mathcal{K} = \mathfrak{F}$ means that ζ lies in $\mathfrak{F}$, as asserted.

5. **Properties of radical extensions.** The radical extensions can be completely characterized by the order $[(R):(F)]$ of the associated divisor group $D = (R)$.

THEOREM 5. *A finite separable extension K/F with radical R and with coefficient field $\mathcal{K}$ is a radical extension of F if and only if*

$$[(R):(F)] = [K:F(\mathcal{K})]. \tag{8}$$

Proof. Suppose first that K is a radical extension so that it is generated by R. The associated group $(R)/(F)$ of divisor classes is then a finite Abelian group so that it may be represented as a direct product of cyclic groups of orders $m_1, \cdots, m_h$. Let the generators of these cycles be the divisors

$$d_1 = (r_1), \cdots, \qquad d_h = (r_h), \qquad \text{order of } r_i = m_i. \tag{9}$$

The exponent of the group $(R)/(F)$ is then

$$e = \text{l. c. m. } (m_1, \cdots, m_h) \tag{10}$$

and the order is the product $m_1 \cdots m_h$. Each generating radical r_i satisfies an equation of the form

$$r_i^{m_i} = b_i x_i \qquad (b_i \text{ in } \mathcal{K},\ x_i \text{ in } F). \tag{11}$$

The radicals $r_1, \cdots, r_h$, together with the elements of F and $\mathcal{K}$, generate the whole group R of radicals; hence the radical extension K is generated by

them as $K = F(\mathcal{K}, r_1, \cdots, r_h)$. In view of the equations (11) for these generating elements, the total degree of this field is at most $[K:F(\mathcal{K})] \leqq m_1 \cdots m_h = [(R):(F)]$. Combined with the reverse inequality as given by (7), this proves the desired result (8).

Incidentally, this argument shows also that each of the equations (11) of degree m_i for a radical r_i of order m_i is an irreducible equation over $F(\mathcal{K})$. This fact could also be established directly, without appeal to Theorem 4.

Conversely, suppose that the relation (8) holds for the radical R of an extension K. The subfield $F(R)$ generated by this radical is then a radical extension which has the same radical R and the same coefficient field $\mathcal{K}$ as does K. Therefore (8) applies to this extension $F(R)$ and gives $[(R):(F)] = [F(R):F(\mathcal{K})]$. By assumption, $[(R):(F)] = [K:F(\mathcal{K})]$. The subfield $F(R)$ has thus the same degree as the whole field so that $F(R) = K$ and K is indeed a radical extension.

As a consequence of this theorem, we can derive a result subsequently useful in recognizing the radicals of certain given extensions.

COROLLARY. *If K is a separable extension of F generated from F by the adjunction of a multiplicative group $R_0 \supset F^*$ which consists of radicals, then $R_0\mathcal{K}^*$ is the radical of K and (R_0) is the group D of all divisors of F principal in K.*

Proof. In any event the group $R_0\mathcal{K}^*$ is contained in the radical R. Suppose first that K/F is finite. The finite group $(R_0)/(F)$ may then be generated exactly as in (9)–(11). The argument used to prove the theorem then shows that $[K:F(\mathcal{K})] = [F(\mathcal{K}, R_0):F(\mathcal{K})] \leqq [(R_0):(F)]$. But (8) then gives

$$[(R):(F)] = [K:F(\mathcal{K})] \leqq [(R_0):(F)].$$

This proves that $(R) = (R_0)$ and hence that $R \subset R_0\mathcal{K}^*$. The group D of principal divisors is $D = (R) = (R_0\mathcal{K}^*) = (R_0)$, as asserted.

If K is infinite over F, the assertions of the corollary may be proved by applying the results of the finite case to a directed system of subgroups $R_\alpha \subset R_0$ of finite order over F^*.

Next, we consider the properties of subfields of a radical extension.

THEOREM 6. *If K is a separable radical extension of F, then any subfield K' of K containing F is also a radical extension.*

For this proof, we denote the radical of any field K relative to a subfield F by $R[K/F]$. We observe at once that the radical of K' is given as the intersection

$$R[K'/F] = R[K/F] \cap K'^*. \tag{12}$$

On the other hand, the composite group $R[K/F]\cdot K'^*$ contains only radicals of K over K'. Since $R[K/F]$ by itself generates the whole field K over F, this larger group $R[K/F]\cdot K'^*$ generates K over K'. By the corollary above, this group is thus the whole radical of K over K' so that[8]

$$R[K/K'] = R[K/F]\cdot K'^*. \tag{13}$$

[8] Incidentally, this formula generalizes Theorem 14 of [6].

By one of the isomorphism theorems for Abelian groups, the two assertions (12) and (13) combine to give the isomorphism

$$R[K/F]/R[K'/F] \cong R[K/K']/K'^{*}. \tag{14}$$

Consider now the special case when K' contains the whole field $\mathfrak{K}$ of coefficients. These relations will then still hold, if in the formulas each group R_0 of radicals is replaced by the corresponding group (R_0) of principal divisors. For example, in (12), a divisor common to the groups $(R[K/F])$ and (K'^{*}) has the form $d = (r) = (u)$, for r a radical of K and u in K'. Then $u = ra$, for some $a \in \mathfrak{K}$. Since $K' \supset \mathfrak{K}$, u also will be a radical of K' so that $d = (u)$ is in the group $(R[K'/F])$. The rest of (12) and (13) are easily established in this case, thus giving also the analogue of (14), for $K' \supset \mathfrak{K}$, as

$$(R[K/F])/(R[K'/F]) \cong (R[K/K'])/(K'^{*}). \tag{14a}$$

Return now to the proof of the theorem in the case when K is finite over F. Consider first the case of an intermediate field K' with $K' \supset \mathfrak{K}$ and suppose, contrary to the result of the theorem, that K' is not a radical extension of F. Thus, Theorems 4 and 5 give

$$[(R[K'/F]):(F)] < [K':F(\mathfrak{K})].$$

Since K still is a radical extension of K', Theorem 5 also gives

$$[(R[K/K']):(K')] = [K:K'].$$

By (14a), the term on the left here may be replaced by the index

$$[(R[K/F]):(R[K'/F])].$$

The two results then combine to give the inequality

$$[(R[K/F]):(F)] < [K:K']\cdot[K':F(\mathfrak{K})] = [K:F(\mathfrak{K})].$$

By Theorem 5, this contradicts the assumption that K is a radical extension of F. Therefore K'/F *is* a radical extension.

Now consider the general case of a field K' not containing the whole constant field $\mathfrak{K}$. It has an extension $L = K'(\mathfrak{K})$ which is a radical extension of F, by the case already treated. Since this field is a coefficient extension of K', Corollary 2 of Theorem 3 asserts that every divisor principal in L is already principal in K' so that $(R[K'/F]) = (R[L/F])$ and

$$[(R[K'/F]):(F)] = [(R[L/F]):(F)] = [L:F(\mathfrak{K})].$$

On the other hand, let $\mathfrak{K}'$ be the field of coefficients of K', and apply Lemma 3 to the finite extension K' of $F(\mathfrak{K}')$. It proves that

$$[L:F(\mathfrak{K})] = [K'(\mathfrak{K}):F(\mathfrak{K})] = [K':F(\mathfrak{K}')].$$

These two results combine to give the equality

$$[(R[K'/F]):(F)] = [K':F(\mathfrak{K}')].$$

By Theorem 5, this suffices to prove K' a radical extension.

To complete the proof of Theorem 6, we need only treat the case of a subfield K' of an infinite extension K. Every element u of K' is contained in a subfield of K generated by a finite number of radicals of K; to this subfield the preceding argument applies to show that the given field K' is indeed a radical extension of F.

COROLLARY 1. *Let D and D' be respectively the groups of divisors of F principal in K and K'. The correspondence $K' \to D'/(F)$ maps the lattice of all those fields K' which contain $\mathfrak{K}$ isomorphically on the lattice of all subgroups of the group $D/(F)$ of divisor classes.*

Proof. It is clear that the correspondence $K' \to D'$ is one which preserves inclusion ($K' \geqq K''$ implies $D' \geqq D''$). To obtain the lattice isomorphism, it then suffices to show that there is one and only one field corresponding to a given group D' of divisors with $D \supset D' \supset (F)$. For given D', let R_0 be the set of all those radicals r of K for which the principal divisor (r) lies in D' and set $K_0 = F(R_0)$. Then R_0 is a group which contains F^* and $\mathfrak{K}^*$ and $(R_0) = D'$ so that the corollary of Theorem 5 shows that K_0 is a radical extension with $K_0 \to D'$. Furthermore, if K' is any other field with $K' \to D'$, this field must contain some one radical r corresponding to each divisor of D' so that it must contain all radicals R_0 used to construct K_0, whence $K' = K_0$, and K_0 is unique. This completes the proof.

COROLLARY 2. *Let R and R' be respectively the radicals of K and K'. The correspondence $K' \to R'$ maps the lattice of all intermediate fields K' isomorphically on the lattice of those subgroups R' of the whole group R which satisfy the conditions* (i) $R' \supset F^*$; (ii) $R' \cap \mathfrak{K}^*$ *is closed under subtraction of distinct elements.*

Proof. Since each K' is a radical extension, it is determined by its radical R'. The correspondence $K' \to R'$ is thus one-one; since it preserves inclusion, it is a lattice isomorphism. It remains only to prove that every group R' satisfying conditions (i) and (ii) is the radical of some intermediate field. Since R' contains -1, it contains $-a$ with a; hence the assumption (ii) will prove $R' \cap \mathfrak{K}^*$ is (except for the absence of 0) a subfield $\mathfrak{P}$ of $\mathfrak{K}$. The field $K' = F(R')$ is then a radical extension of F and its radical may be computed, by the corollary to Theorem 5, to be just $R'\mathfrak{K}'^*$, where $\mathfrak{K}'$ is the field of constants from K'. If we can prove $\mathfrak{K}' = \mathfrak{P}$, we shall have proved that the given group R' is the radical of a field K'.

Suppose instead that $\mathfrak{K}' > \mathfrak{P}$. Then $F(\mathfrak{K}') > F(\mathfrak{P})$ so that Theorem 5 gives

$$[K':F(\mathfrak{P})] > [K':F(\mathfrak{K}')] = [(R'):(F)].$$

Now the correspondence $r \to (r)$ maps the Abelian group $R'/F^*\mathfrak{L}^*$ isomorphically on $(R')/(F)$. As in (9), one may choose a basis $(r_1), \cdots, (r_h)$ for the latter Abelian group; the group R' is then generated over $F^*\mathfrak{L}^*$ by the representatives $r_1, \cdots, r_h$ with $r_i^{m_i} = b_i x_i$, where each b_i is in $\mathfrak{L}$ and each x_i in F, while $m_1 m_2 \cdots m_h = [(R'):(F)]$. The field $K' = F(\mathfrak{L}, R') = F(\mathfrak{L}, r_1, \cdots, r_h)$ then has a degree at most

$$[K':F(\mathfrak{L})] \leqq m_1 m_2 \cdots m_h = [(R'):(F)].$$

This contradicts the previous inequality and hence the assumption $\mathfrak{K}' > \mathfrak{L}$.

A radical extension is not uniquely determined up to isomorphism by its divisor group and its field $\mathfrak{K}$ of coefficients, as one may readily show by examples. Instead, one has the following modified uniqueness theorem, in which, for simplicity, all extensions of F are supposed to be within a fixed algebraically complete extension of F.

THEOREM 7. *If K and K' are separable radical extensions of F with the same group D of associated divisors, then there exists an algebraic extension $\mathfrak{W}$ of the field of coefficients such that $K(\mathfrak{W}) = K'(\mathfrak{W})$. In case K is finite over F, one may also require that $\mathfrak{W}$ be finite over $\mathfrak{F}$.*

Proof. Suppose first that K/F is finite, and generate K as $K = F(\mathfrak{K}, r_1, \cdots, r_h)$ by radicals r_i, as in (9), (10), and (11). With the same generation of the group of divisors, the second radical extension K' will have a generation $K' = F(\mathfrak{K}, r_1', \cdots, r_h')$ in which the r_i' satisfy equations $r_i'^{m_i} = b_i' x_i$, of the form of equations (11). The field $\mathfrak{W}$ obtained from $\mathfrak{F}$ by the adjunction of $\mathfrak{K}$, $\mathfrak{K}'$, and *all* the roots of the equations $t_i^{m_i} = b_i/b_i'$ then has the desired properties. The case of an infinite field K is treated by the same argument applied to the finite subfields of K.

6. **The group of radicals.** The ordinary Kummer theory studies extensions K of a field F which can be generated by the adjunction of n-th roots to F (n fixed and prime to the characteristic). On the assumption that all n-th roots of unity lie in F, each such Kummer field K is uniquely determined by a certain multiplicative group in F; namely, the group of all those elements of F which have n-th roots in K (see, e.g., [20], [3]). In similar vein, the class field theory determines all Abelian extensions of number fields and of certain function fields uniquely in terms of suitable multiplicative groups. If one seeks for a simple multiplicative group which is related in similar manner to the radical extensions studied here, the answer is to be found, not in the group of divisors (Theorem 7), but in the group R of all radicals.

For a given separable extension K, the group R of radicals is an Abelian group with the following properties:

(i) R contains the multiplicative groups F^* and $\mathfrak{K}^*$ of the fields F and $\mathfrak{K}$, and thus also the group composite $F^*\mathfrak{K}^*$;

(ii) $R/\mathfrak{K}^*$ has no elements of finite order;

(iii) $R/F^*\mathcal{K}^*$ has every element of a finite order, prime to the characteristic of F.

Abstractly speaking, R is not merely a multiplicative group; it is rather a group of which certain portions are the multiplicative groups of fields. (This corresponds to the fact that the radical extensions include all possible coefficient extensions so that any tool adequate to describe all radical extensions must describe all algebraic extensions of the constant field $\mathfrak{F}$.) However, instead of treating R postulationally, we now show why the structure of a group R with these properties does determine the structure of K.

THEOREM 8. *Let K and K' be separable radical extensions of F with radicals R and R' which are isomorphic under a correspondence $r \leftrightarrow r^\phi$ which is the identity on the subgroup F^* and which carries the coefficient field $\mathcal{K}$ isomorphically onto $\mathcal{K}'$. Then ϕ may be extended in one and only one way to an isomorphism of K to K'.*

Observe that the hypotheses placed on the correspondence ϕ in this theorem refer only to the structure of R relative to the subsystems F^*, $\mathcal{K}^*$, as summarized above.

Proof. Assume first that $(R)/(F^*)$ is finite. By the assumption on the coefficient field, the given isomorphism ϕ can be extended to an isomorphism of $F(\mathcal{K})$ to $F(\mathcal{K}')$. The field K' can be generated from $F(\mathcal{K})$ by the successive adjunction of the roots r_i of the irreducible equations $t_i^{m_i} = b_i x_i$, as in (11). These equations are purely multiplicative, so the corresponding generating radicals r_i' of K' will satisfy corresponding irreducible separable equations $t_i^{m_i} = b_i^\phi x_i$. Repeated applications of a basic extension theorem of the Galois theory then give one and only one isomorphism which extends ϕ and maps r_i on r_i'. This completes the finite case.

The infinite case is treated by the usual approximation devices. For each subgroup $R_0 \subset R$ with a finite order over $F^*\mathcal{K}^*$, one has a unique extension of ϕ to an isomorphism of $F(R_0)$ to $F(R_0')$. Because of the uniqueness of the extension, two such extensions agree wherever their fields overlap. They may then be combined to give a single extension to all of K.

COROLLARY. *The group of all automorphisms T of a separable radical extension K/F with radical R is the group of all those (multiplicative) automorphisms $r \leftrightarrow r^T$ of the group R which have the following properties*: (i) $x^T = x$ *if* $x \in F^*$; (ii) $a^T \in \mathcal{K}^*$ *if and only if* $a \in \mathcal{K}^*$; (iii) $(a + b)^T = a^T + b^T$, *provided* a, b, *and* $a + b$ *are all in* $\mathcal{K}^*$.

Proof. Every field automorphism induces such a T; conversely, each such T will by the theorem give an automorphism of K, provided T is an automorphism of the field $\mathcal{K}$. This follows from the conditions (ii) and (iii), with the observation that $(-a)^T = [(-1)a]^T = (-1)a^T = -a^T$, so that (iii) holds even if $a + b = 0$.

A companion to the above uniqueness theorem is the following existence theorem.

THEOREM 9. *If $\mathcal{K}$ is any separable algebraic extension of the coefficient field $\mathfrak{F}$ of F and if $R \supset \mathcal{K}^*$ is an Abelian group with the properties* (i), (ii), (iii) *above* (*pp.* 142–143), *then there exists one* (*and essentially only one*) *separable radical extension K of F which has $\mathcal{K}$ as field of coefficients and R as radical.*

Observe that we assume R given with a *specified* subgroup $\mathcal{K}^*$; it would be *a priori* possible that more than one subgroup of a given R could be interpreted as the multiplicative group of a field. The theorem itself is essentially just an expression of the fact that the equations (11) which define a radical extension over its coefficient extension are properly multiplicative in form so they are determined by the group structure of R.

Proof. Since every element of $R/F^*\mathcal{K}^*$ has finite order, it will suffice, as in previous cases, to suppose that $R/F^*\mathcal{K}^*$ is finite. Because of assumptions (ii) and (iii), the group $R/\mathcal{K}^* = (R)$ can be identified with a group D of fractional divisors of F. Let (r) denote the divisor belonging to the element $r \in R$. Select a basis

$$d_1 = (r_1), \cdots, \qquad d_h = (r_h), \qquad \text{order of } r_i = m_i, \tag{15}$$

for the finite Abelian group $R/F^*\mathcal{K}^*$. Then $t = r_i$ must satisfy a separable equation of the form

$$t^{m_i} = x_i b_i, \qquad b_i \in \mathcal{K},\ x_i \in F. \tag{16}$$

Using any formal roots r_i' of these h equations over $F(\mathcal{K})$, one can now construct a field K' as $K' = F(\mathcal{K}, r_1', \cdots, r_h')$. By means of the previously established properties of radical extensions, one then sees that this field is indeed a radical extension, that its divisor group D' is exactly the previous group D, that its radical R' is isomorphic to the given radical R under a correspondence $r_i \to r_i'$ and that its field of coefficients is just $\mathcal{K}$. We have thus found a field with a given radical R; the uniqueness of the construction is asserted by the previous theorem.

As another illustration of the adequacy of the multiplicative description of R by (i), (ii), and (iii), consider three fields $F \subset L \subset K$. Given the radicals $R[K/L]$ and $R[L/F]$ as groups, the radical $R[K/F]$ is completely determined as the set of all those elements r in $R[K/L]$ for which some power r^m is in the subgroup $F^*\mathcal{K}^*$ of $R[K/L]$. In terms of the subgroup L^* of $R[K/L]$ one may state

THEOREM 10. *If K is a separable radical extension of L and L a separable radical extension of F, then K is a radical extension of F if and only if $R[K/L]$ is the group composite $R[K/F] \cdot L^*$.*

Proof. The necessity of this condition was established in (13) of §5. To prove the sufficiency, observe that $R[K/F] \supset R[L/F]$ so that the adjunction to F of the radical $R[K/F]$ will first generate the whole field L, then the whole radical $R[K/F] \cdot L^* = R[K/L]$, and therefore the whole of the radical extension K/L.

Arithmetically, one of the most important types of function fields is that in which the coefficient field $\mathfrak{F}$ is a *finite* field, say with $q = p^n$ elements. In this case, the description of the group R can be substantially simplified because the part of R to be identified with the coefficient extension $\mathcal{K}$ can be described in purely group-theoretic terms. Indeed, a finite extension $\mathcal{K}$ of degree f over $\mathfrak{F}$ will have q^f elements and its multiplicative group is cyclic of order $q^f - 1$.

Conversely, let $\mathfrak{H}$ be a cyclic multiplicative group of order $q^f - 1$ with $\mathfrak{F}^*$ as a subgroup. We assert that $\mathfrak{H}$ can be mapped isomorphically onto the multiplicative group $\mathcal{K}^*$ of the finite extension $\mathcal{K}$ of degree f over $\mathfrak{F}$ in such manner that the elements of $\mathfrak{F}^*$ are all left fixed in the mapping (without this last proviso, the assertion would be trivially true). For, let ω be any element generating the multiplicative group $\mathcal{K}^*$, so that

$$\omega^g = \zeta, \qquad g = (q^f - 1)/(q - 1),$$

is an element generating the multiplicative group $\mathfrak{F}^*$. If σ is any generator of the given group $\mathfrak{H}$, then

$$\sigma^g = \zeta^i, \qquad (i, q - 1) = 1.$$

Any other generator σ' of $\mathfrak{H}$ would have the form $\sigma' = \sigma^y$, for some integer y prime to $q^f - 1$ and for this generator $(\sigma')^g = \sigma^{yg} = \zeta^{iy}$. Let $l_1, \cdots, l_t$ be the various prime factors of $q^f - 1$ which are not divisors of $q - 1$. Then the congruences

$$iy \equiv 1 \pmod{q - 1}, \qquad y \equiv 1 \pmod{l_1}, \qquad \cdots, \qquad y \equiv 1 \pmod{l_t}$$

have a solution y which will be relatively prime to $q^f - 1$ because it is prime to the factors l_i and to the factor $q - 1$. This solution y determines a generator σ' for which $(\sigma')^g = \zeta^{iy} = \zeta$. The correspondence $\sigma' \to \omega$ then maps $\mathfrak{H}$ isomorphically upon $\mathcal{K}^*$, and leaves ζ and its powers fixed, as was asserted.

In view of this construction, Theorem 9 now becomes the following assertion.

THEOREM 11. *Let F be a field of characteristic p with finite subfield $\mathfrak{F}$, of q elements, which is algebraically closed in F. Then an Abelian multiplicative group R is the radical of a separable extension K of F if and only if*

(i) *R contains the multiplicative group F^* of F.*

(ii) *The elements of finite order in R form a cyclic group of order $q^f - 1$ for some integer f.*

(iii) *R/F^* has every element of finite order.*

Observe, however, that we no longer have an analogue of Theorem 8 (the uniqueness theorem) for a radical as described merely by the properties (i), (ii), (iii) of this theorem; in other words, the requirement of Theorem 8 that the mapping ϕ of the radical R upon another radical R' be an isomorphism of the *field* $\mathcal{K}$ is essential. This is because the above interpretation of the multiplicative group $\mathfrak{H}$ as the group $\mathcal{K}^*$ of a field can usually be performed in many different

ways, which are not all equivalent under isomorphisms of $\mathfrak{K}$. This is the case, for instance, if $q = 3, f = 2$.

CHAPTER II

GALOIS THEORY OF RADICAL EXTENSIONS

7. **Conditions for normality.** For a given radical extension K/F every automorphism S, with $u \leftrightarrow u^S$, induces an automorphism σ of the corresponding extension $\mathfrak{K}/\mathfrak{F}$. Furthermore, each radical r must be mapped upon another radical r^S with the same divisor so that S determines a set of constants $a = a(S, r)$ with

$$r^S = a(S, r) \cdot r, \qquad \text{where } a(S, r) \in \mathfrak{K}^*. \tag{1}$$

The product ST of two automorphisms will mean: first apply T, then S, so that $r^{(ST)} = (r^T)^S$. One then has

$$a(S, r)a(T, r)^\sigma = a(ST, r); \tag{2}$$

while, for the product of two radicals,

$$a(S, rr') = a(S, r)a(S, r'). \tag{3}$$

If r is a radical of order m, with $r^m = bx$, for b in $\mathfrak{K}$ and x in F, the application of S to this equation proves that

$$[a(S, r)]^m = b^{\sigma-1}, \qquad (r^m = bx). \tag{4}$$

In particular, if r is in F, then $a(S, r) = 1$.

The Galois group of a normal extension K/F will be analyzed in terms of the function $a(S, r)$ with these properties (2), (3), and (4). If K is a finite radical extension, generated by h radicals r_i, as in (9), (10), and (11) of §5, then the function $a(S, r)$ is essentially determined by the h partial functions

$$a_i(S) = a(S, r_i), \qquad i = 1, \cdots, h. \tag{5}$$

From these, the other values of $a(S, r)$ may be computed, using (3) and (4). In the presence of the necessary condition (4) these functions determine the automorphism in the following sense.

LEMMA 5. *Let K/F be a separable radical extension generated by radicals $r_1, \cdots, r_h$ of orders $m_1, \cdots, m_h$, as in §5; let σ be an automorphism of the coefficient extension $\mathfrak{K}/\mathfrak{F}$; and let the constants $a_1, \cdots, a_h$ in $\mathfrak{K}$ satisfy the conditions*

$$a_i^{m_i} = b_i^{\sigma-1}, \qquad i = 1, \cdots, h. \tag{6}$$

Then there exists one and only one extension S of σ which is an automorphism of K over F with the property

$$r_i^S = a_i r_i, \qquad i = 1, \cdots, h. \tag{7}$$

Proof. Any radical r of K has the form

$$r = r_1^{e_1} \cdots r_h^{e_h} cy, \qquad 0 \leqq e_i < m_i, \qquad c \in \mathcal{K}, y \in F. \tag{8}$$

We define a corresponding radical r^S by

$$r^S = a_1^{e_1} \cdots a_h^{e_h} c^{\sigma-1} r. \tag{9}$$

The assumed condition (6) insures that this formula holds for any exponents e_i, even if $e_i \geqq m_i$. Thus (9) gives an automorphism $r \leftrightarrow r^S$ of R to itself which realizes the prescription (7). Theorem 8 then gives the extension of this group automorphism to one of the field K.

We next derive a condition for normality which bears a close resemblance to a theorem due to Albert, giving conditions that an extension of a p-adic field be normal ([1], Theorem 6).

THEOREM 12. *A finite separable radical extension K/F is normal if and only if the following conditions hold*:

(i) *The coefficient extension $\mathcal{K}/\mathfrak{F}$ is normal*;

(ii) *$\mathcal{K}$ contains all e-th roots of unity, where e is the exponent of the associated group $(R)/(F)$ of divisor classes*;

(iii) *For each automorphism σ of $\mathcal{K}/\mathfrak{F}$ and for each radical r of order m with $r^m = bx$, for b in $\mathcal{K}$, x in $\mathfrak{F}$, the constant $b^{\sigma-1}$ is an m-th power in $\mathcal{K}^*$.*

Remarks. Essentially the same theorem holds for an infinite extension, if (ii) is appropriately modified to require that all m-th roots of unity be in $\mathcal{K}$ for each order m of a divisor of $(R)/(F)$; alternatively, one may still use condition (ii) as stated, if the exponent e be interpreted as a "Steinitz G number" (the formal l.c.m. of the orders of all elements of $(R)/(F)$, written as a possibly infinite product of primes with possibly infinite exponents).

If a finite extension K/F is generated over $F(\mathcal{K})$ by radicals $r_1, \cdots, r_h$ as in §5, the condition (iii) of this theorem may be replaced by the parallel condition applied to these radicals alone:

(iiia) For each σ, $b_i^{\sigma-1}$ is an m_i-th power in $\mathcal{K}^*$, for $i = 1, 2, \cdots, h$.

Proof. Assume first that K/F is normal. The condition (i) is immediate, and (iii) follows from the equation (4) deduced above. Finally, if e is the exponent of $(R)/(F)$, there is a radical r of order $m = e$ which satisfies over $F(\mathcal{K})$ an equation $t^e = bx$ which is irreducible over $F(\mathcal{K})$. Since one root r of this equation lies in K, the other roots $\zeta^i r$ also lie in K, where ζ is a primitive e-th root of unity. By the assumption of separability, e is prime to the characteristic.

Conversely, it is clear that condition (iii) implies its special case (iiia), so that it will suffice to show that (i), (ii), and (iiia) insure the normality of the extension $K = F(\mathcal{K}, r_1, \cdots, r_h)$. Let $r = r_i$ be one of these generating radicals, satisfying an equation $t^m = bx$ over $F(\mathcal{K})$, for $b = b_i$, $m = m_i$. Over F this radical satisfies the equation

$$f(t) = \prod_{\sigma} (t^m - b^\sigma x).$$

By assumption, $b^{\sigma-1}$ has the form $(a_\sigma)^m$ for some constant a_σ in $\mathcal{K}$, and $\mathcal{K}$ contains a primitive m-th root of unity ζ. One then computes that the quantities $\zeta^j a_\sigma r$, for $j = 0, 1, \cdots, m-1$, include all the roots of the polynomial $f(t)$. They all lie in the field K, so that K, as generated by the normal extension $\mathcal{K}$ and the roots r_i of these polynomials $f(t)$, is itself normal over F, as asserted in the theorem.

8. **Characters and the Galois group.** Let H be the Galois group of a separable normal radical extension K over its coefficient extension $F(\mathcal{K})$. This subgroup H of the whole Galois group depends essentially on the (ordinary) characters of the divisor class group $D = (R)/(F)$. Such a character is essentially a function C which defines for each radical r in R a value $C(r)$ in the coefficient field $\mathcal{K}$ such that

$$C(rr') = C(r)C(r'), \tag{10}$$

$$(r) \in (F) \quad \text{implies} \quad C(r) = 1. \tag{11}$$

By the last theorem, the presence of a radical r of order m in K insures the presence in $\mathcal{K}$ of all m-th roots of unity; hence every linear character which could be realized in an algebraically complete field (of the same characteristic as $\mathcal{K}$) can already be realized in $\mathcal{K}$. Two characters C_1, C_2 have a product C_1C_2 defined by

$$C_1C_2(r) = C_1(r)C_2(r), \qquad r \in R. \tag{12}$$

Under this multiplication, the characters C form a group X; in case D is finite, X is isomorphic to D.

THEOREM 13.[9] *If a radical extension K is normal and separable over F, then each character C of its associated divisor class group $(R)/(F)$ yields an automorphism $u \leftrightarrow u^S$ of K over $F(\mathcal{K})$, determined by the formulas*

$$r^S = C(r)\cdot r, \qquad \text{for each } r \in R. \tag{13}$$

This correspondence of characters to automorphisms is an isomorphism of the character group X to the whole Galois group of K over the coefficient extension $F(\mathcal{K})$.

Remark. In the case of a finite extension K/F, the conclusion of the theorem means that the Galois group of $K/F(\mathcal{K})$ is isomorphic to the divisor group, regarded as (isomorphic to) its own character group; for the infinite case the character group is more useful.

Proof. Let S be a fixed automorphism of $K/F(\mathcal{K})$. According to (3) the function $a(S, r)$ is a homomorphism of R to the multiplicative group of $\mathcal{K}^*$. By (4), any radical r of order 1 is mapped onto 1, for in the present case the induced automorphism σ of $\mathcal{K}$ is the identity. Therefore the function can be interpreted as a character $C_S(r) = a(S, r)$. According to (2), with $\sigma = 1$, the

[9] This theorem, in a form using characters, was first proved by Baer. His statement is given by the equivalence of the properties (1) and (3) of his Theorem 7.1, in [3.]

product of automorphisms corresponds to the product of characters. Since the whole field K is generated by radicals, distinct isomorphisms must have different effects on at least one radical. We conclude that the correspondence $S \to C_S$ does map the automorphism group of $K/F(\mathfrak{K})$ isomorphically on *part* of the character group C, and that the relation (13) does hold for these automorphisms.

It remains only to show that every character can be realized by such an automorphism. If K is finite, this is immediate, for the number of characters equals the order of $(R)/(F^*)$, which in turn (Theorem 5) equals the degree of $K/F(\mathfrak{K})$. If K is infinite, it may be represented as the join of the set of all its subfields K_ν which are finite over $F(\mathfrak{K})$. Each of these subfields is a radical extension (Theorem 6), hence by Theorem 12 is normal over F. Now let C be any character of R; it clearly induces a character C_ν on the radical R_ν of K_ν, for every ν. By the finite case, each such character C_ν gives one and only one automorphism S_ν of $K_\nu/F(\mathfrak{K})$. If $K_\nu \subset K_\mu$, the uniqueness of the automorphism S_ν shows that S_μ must be a prolongation of S_ν. The family of automorphisms S_ν may therefore be combined to give a single automorphism S of K, which agrees with each S_ν on the corresponding subfield K_ν. The map $S \to C_S$ of this composite automorphism S is exactly the given character. This completes the proof of the theorem.

COROLLARY. *If K is an infinite extension, the correspondence of characters to automorphisms is continuous in both directions.*

Proof. X is the character group of a discrete Abelian group $(R)/(F)$ of divisor classes. Its topology is usually defined by the following complete system of neighborhoods of unity. Each finite set R_0 of radicals determines a neighborhood $N(R_0)$ consisting of all those characters C for which

$$C(r_0) = 1 \qquad \text{whenever } r_0 \in R_0 .$$

Because every divisor class has finite order, one obtains an equivalent complete system of neighborhoods of 1 if one restricts R_0 to be a *subgroup* of finite order over $F^*\mathfrak{K}^*$. On the other hand, each subfield K_0 of finite degree over $F(\mathfrak{K})$ defines a neighborhood of unity in the Galois group, consisting of all those automorphisms S for which

$$u^S = u \qquad \text{whenever } u \in K_0 .$$

But we know that the finite groups R_0 of radicals correspond to the finite subfields $K_0 = F(\mathfrak{K}, R_0)$; it follows that the correspondence $S \to C_S$ maps one system of neighborhoods on the other, and hence is indeed bicontinuous.

9. **Analysis of the Galois group.** Any separable normal extension K of a function field F decomposes naturally into two stages: from F to $F(\mathfrak{K})$ to K. If H is the Galois group of $K/F(\mathfrak{K})$ and Γ the Galois group of $\mathfrak{K}/\mathfrak{F}$, the whole Galois group G of K/F is a group extension of H by Γ, in the sense that G has H as a normal subgroup and $G/H \cong \Gamma$ as the corresponding factor group. Such

an extension may be described by choosing in G for each automorphism σ of Γ a representative U_σ which induces σ. Every element of G then has the form SU_σ, for some $S \in H$, $\sigma \in \Gamma$. If H is Abelian, the multiplication of these elements is determined by a table

$$U_\sigma T U_\sigma^{-1} = T^\sigma \qquad (T \in H, \sigma \in \Gamma), \tag{14}$$

$$U_\sigma U_\tau = S_{\sigma,\tau} U_{\sigma\tau} \qquad (\text{each } S_{\sigma,\tau} \in H). \tag{15}$$

Here (14) indicates that transformation by each σ determines an automorphism $T \to T^\sigma$ of H; it is independent of the choice of U and has $(T^\sigma)^\tau = T^{\tau\sigma}$. In (15) $S_{\sigma,\tau}$ is a *factor set*[10] of Γ in H. We wish to find the special conditions holding for these automorphisms and factor sets in the case when K is a radical extension.

THEOREM 14. *If K is a normal radical extension of F, then each character $C(S)$ of the Galois group H of $K/F(\mathcal{K})$ into the field $\mathcal{K}$ has the following properties*

$$C(S^\sigma) = [C(S)]^\sigma, \qquad \textit{for each } S \, S \in H \textit{ and } \sigma \in \Gamma; \tag{I}$$

$$C(S_{\sigma,\tau}) \sim 1 \qquad \textit{in}^{11}\ \mathcal{K}. \tag{II}$$

Proof. We already know that the function $a(S, r)$ of (1) for fixed S in H is a character of $(R)/(F)$, and that the various automorphisms S in H give all such characters. Therefore, for fixed r, $C_r(S) = a(S, r)$ is a character of H in $\mathcal{K}$, and all characters have this form, by the duality between a discrete Abelian group and its character group. Repeated applications of (2) now give

$$\begin{aligned} a(U_\sigma S U_\sigma^{-1}, r) &= a(U_\sigma, r) a(S, r)^\sigma a(U_\sigma^{-1}, r)^\sigma \\ &= a(U_\sigma U_\sigma^{-1}, r) a(S, r)^\sigma = a(S, r)^\sigma. \end{aligned}$$

This proves (I). To obtain (II), apply (2) to both sides of the equation $a(U_\sigma U_\tau, r) = a(S_{\sigma,\tau} U_{\sigma\tau}, r)$. The result is

$$a(S_{\sigma,\tau}, r) = a(U_\sigma, r) a(U_\tau, r)^\sigma / a(U_{\sigma\tau}, r).$$

The factor set on the right is indeed similar to 1 in $\mathcal{K}$.

The condition (II) may be put into a number of different but essentially equivalent forms. If the Abelian group H is finite, and has a basis of automorphisms $S_1, \cdots, S_h$ of orders $m_1, \cdots, m_h$, then the group of characters is generated by the characters C_i with

$$C_i(S_i) = \zeta_i, \qquad C_i(S_j) = 1, \qquad i \neq j,$$

where ζ_i is a primitive m_i-th root of unity in $\mathcal{K}$. The condition then has the form

$$C_i(S_{\sigma,\tau}) \sim 1 \qquad \text{in } \mathcal{K},\ i = 1, \cdots, h. \tag{IIa}$$

[10] For the theory of factor sets, see, for example, [21], or [12].

[11] $f_{\sigma,\tau} \sim 1$ means that the factor set $f_{\sigma,\tau}$ of elements in $\mathcal{K}$ is similar to 1; i.e., that there exist constants a_σ in $\mathcal{K}^*$ such that $f_{\sigma,\tau} = a_\sigma a_\tau^\sigma / a_{\sigma\tau}$.

In terms of a set of generating radicals $r_1, \cdots, r_h$ this may also be stated as

(IIb) $$a(S_{\sigma,\tau}, r_i) \sim 1 \qquad \text{in } \mathcal{K},\ i = 1, \cdots, h.$$

If each automorphism is expressed in terms of the generators, the factor set becomes

$$S_{\sigma,\tau} = S_1^{f_{1,\sigma,\tau}} S_2^{f_{2,\sigma,\tau}} \cdots S_h^{f_{h,\sigma,\tau}},$$

where, for each $i = 1, \cdots, h$, $f_{i,\sigma,\tau}$ is a factor set of integers modulo m_i, relative to suitable automorphisms induced by Γ on these integers. The condition (II) then takes the form

(IIc) $$[C_i(S_i)]^{f_{i,\sigma,\tau}} \sim 1 \qquad \text{in } \mathcal{K},\ i = 1, \cdots, h.$$

One may also observe that the conditions (I) and (II) do not depend essentially upon the choice of the representatives U_σ used in (14) and (15) to describe the group G; the validity of (I) and (II) for any one set of such representatives implies the validity for any other set.

The main result of our analysis is the conclusion that the conditions (I) and (II) are not only necessary, but also sufficient for the realization of such a group extension by a field generated by radicals. This result may be stated in full as follows, for given F and $\mathfrak{F}$.

THEOREM 15. *Let a finite group G be represented as in* (14) *and* (15) *as an extension of its normal subgroup H by the factor group $\Gamma = G/H$, where*

(i) *Γ is the Galois group of a finite separable normal extension $\mathcal{K}$ of $\mathfrak{F}$,*

(ii) *H is an Abelian group with exponent e prime to the characteristic of $\mathfrak{F}$,*

(iii) *$\mathcal{K}$ contains all e-th roots of unity,*

(iv) *The group X of characters of H is isomorphic to a group $D/(F)$ of divisor classes from F.*

Then G can be realized as the Galois group of a normal separable radical extension K/F with coefficient field $\mathcal{K}$ if and only if each character C of H in $\mathcal{K}$ satisfies the conditions (I) *and* (II) *of Theorem* 14. *When these conditions hold, a field K may be so constructed that D is the group of all divisors which become principal in K.*

Remark. One may state a companion theorem in which $\mathfrak{F}$ but not F is given. Since one may construct a field F which realizes any specified group of fractional divisor classes, say by making F a field $\mathfrak{F}(x_1, \cdots, x_n)$ of rational functions[12] of n indeterminates, this theorem would differ from the present one only in the omission of hypothesis (iv). However, for an arbitrary F, hypothesis (iv) is necessary.

Proof. The conditions (I) and (II) are already known to be necessary; hence we need only assume them valid and construct a corresponding field. Let the divisors $d_1, \cdots, d_h$ of orders $m_1, \cdots, m_h$ be a basis of the finite Abelian group $D/(F)$, so that $d_i^{m_i} = (x_i)$ for elements x_i in F. Because of (iv), one may

[12] Observe that we consider here the group of (abstract) fractional divisors, not the group of arithmetic divisors.

interpret each divisor d of D as a character $d(S)$ of H in $\mathcal{K}$. According to condition (II), there are elements $a_i(\sigma)$ in $\mathcal{K}$ such that

$$d_i(S_{\sigma,\tau}) = a_i(\sigma)a_i(\tau)^{\sigma}/a_i(\sigma\tau). \tag{16}$$

Because d_i is an element of order m_i in $D/(F)$, the m_i-th power of this equation gives

$$1 = d_i^{m_i}(S_{\sigma,\tau}) = a_i(\sigma)^{m_i}a_i(\tau)^{m_i\sigma}/a_i(\sigma\tau)^{m_i}.$$

This asserts that the function $f_i(\sigma) = [a_i(\sigma)]^{m_i}$ is a crossed character of the Galois group Γ in the normal extension $\mathcal{K}/\mathfrak{F}$. Noether's principal genus theorem applies to this case, and states that the crossed character is a unit character, so that there exist constants b_i in $\mathcal{K}^*$ for which

$$a_i(\sigma)^{m_i} = b_i^{\sigma-1} \qquad \text{(all } \sigma \text{ in } \Gamma\text{)}. \tag{17}$$

These elements b_i are used to construct the field, just as if they were the constants of the analogous equations (4). Observe that each divisor d_i of order m_i has $d_i^{m_i} = (x_i)$ for some element x_i in F. Construct K by adjoining to $F(\mathcal{K})$ roots $r_1, \cdots, r_h$ of the h equations[13]

$$r_i^{m_i} = b_i x_i, \qquad i = 1, 2, \cdots, h.$$

The field $K = F(\mathcal{K}, r_1, \cdots, r_h)$ so constructed is a radical extension with the given group D as its associated group of divisors (Theorem 5, Corollary). The choice of b_i in (17) insures that it is normal (Theorem 12).

The Galois group G' of K/F is described as in Theorem 14. Each automorphism S of the given group H may be interpreted as a character $d \to d(S)$ of the class group $D/(F)$, and hence as an automorphism S' of $K/F(\mathcal{K})$, with

$$r^{S'} = d(S)r, \qquad d = (r), \tag{18}$$

just as in Theorem 13. We may identify S with S' and therefore H with the Galois group of $K/F(\mathcal{K})$. The factor group $\Gamma' = G'/H$ is the group of $\mathcal{K}/\mathfrak{F}$, so may be identified with the given finite group Γ.

Further identification depends on the automorphisms and the factor sets. For this, we must first select in G' appropriate representatives U' of the automorphisms σ of Γ. Because of the choice (17) of the b_i, the construction of Lemma 5 gives such automorphisms U'_σ with

$$r_i^{U'_\sigma} = a_i(\sigma)r_i, \qquad i = 1, \cdots, h. \tag{19}$$

Computation with (18) and (19) gives

$$r_i^{U'_\sigma S} = [d_i(S)r_i]^{U'_\sigma} = [d_i(S)]^{\sigma}a_i(\sigma)r_i,$$

$$r_i^{S^\sigma U'_\sigma} = [a_i(\sigma)r_i]^{S^\sigma} = [d_i(S^\sigma)]a_i(\sigma)r_i.$$

[13] This construction is not uniquely determined, because of the many possible choices for the elements x_i.

These results are equal by condition (I), applied with $C = d_i$; hence

$$U'_\sigma S = S^\sigma U'_\sigma \qquad (S \text{ in } H,\ \sigma \text{ in } \Gamma). \tag{20}$$

A similar computation shows that

$$U'_\sigma U'_\tau = S_{\sigma,\tau} U'_{\sigma\tau}\,. \tag{21}$$

The results (20) and (21) mean that the multiplication table for G' is the same as the table (14) and (15) for G. Hence the group G' of the radical extension is essentially the given group G.

COROLLARY 1. *If D is a divisor group of finite exponent e in F, with e prime to the characteristic of F, then there is a separable Abelian extension of F in which every divisor of D becomes principal if and only if all e-th roots of unity lie in F. When this condition is satisfied, one can find a normal radical extension of F which has no coefficient extension, and in which all divisors of D become principal.*

This corollary emphasizes the extent to which the theory of radical extensions is *not* a theory of Abelian extensions, although in the classical case of algebraic number fields the study of the principal ideal theorem centers around Abelian class fields.

Proof. If D becomes principal in a normal extension K, it already becomes principal in some radical extension $K' \subset K$. If there is some e-th root of unity not present in F, this root appears at least once as the value of a character $C(r)$ for K'. The equations (14) and (I) then show that the Galois group of K' is not Abelian. Hence our condition is necessary.

Conversely, assume all e-th roots present, and apply Theorem 15 with $\mathfrak{K} = \mathfrak{F}$. The construction of this theorem does yield a radical extension in which D becomes principal; in the present case the assumption $\mathfrak{K} = \mathfrak{F}$ insures that the Galois group is the group X of characters, hence is Abelian.

COROLLARY 2. *If D is a finite divisor group of exponent e in F, with e prime to the characteristic, then there exists a finite normal extension K/F in which every divisor of D becomes principal, and in which the coefficient field $\mathfrak{K}$ of K is obtained from $\mathfrak{F}$ by adjunction of all e-th roots of unity. Furthermore, the Galois group of K/F is metabelian.*

Proof. This is an immediate corollary of the preceding construction.

10. **Computations for crossed characters.** Our next aim is to show that any normal extension K with a group which satisfies the essential conditions (I) and (II) is necessarily a radical extension. To this end, we use the decomposition of the group G into H and $\Gamma = G/H$ to study the crossed characters introduced in §4.

Let $c_S = c(S)$ be a crossed character of G in $\mathfrak{K}$. Any element of G has the form SU_σ, for S in H, σ in Γ; while by the definition of a crossed character

$$c(SU_\sigma) = c(S)c(U_\sigma) \qquad (S \text{ in } H). \tag{22}$$

Hence c is completely determined by its values in the subgroup H and by the values of the auxiliary function $c_\sigma = c(U_\sigma)$. One shows at once that these functions satisfy the conditions

$$c(S)c(T) = c(ST) \qquad (S, T \text{ in } H), \tag{23}$$

$$[c(S)]^\tau = c(S^\tau) \qquad (S \text{ in } H), \tag{24}$$

$$c_\sigma c_\tau^\sigma = c(S_{\sigma,\tau})c_{\sigma\tau}\,. \tag{25}$$

Conversely, given functions $c(S)$ and c_σ, for S in H, which satisfy these three conditions, one computes at once that the formula (22) does give a crossed character, according to definition. The crossed characters therefore correspond to the pairs of solutions $(c(S), c_\sigma)$ of (23)–(25). The unit characters correspond to solutions with $c(S) = 1$, $c_\sigma = b^{1-\sigma}$, for some b in $\mathcal{K}$.

Suppose now that $c(S)$ for S in H is a solution of the first two equations (23) and (24). If the final equation (25) has, for given $c(S_{\sigma,\tau})$, two solutions c_σ and c_σ', then their quotient c_σ/c_σ' is itself a crossed character of Γ in $\mathcal{K}$. By the principal genus theorem, this quotient has the form $c_\sigma/c_\sigma' = b^{1-\sigma}$, so that $c_\sigma = b^{1-\sigma}c_\sigma'$. This means that two different solutions of (25) simply give associate crossed characters of G. Therefore classes of crossed characters of G in $\mathcal{K}$ correspond to those distinct solutions of (23) and (24) for which (25) can be solved. But the requirement that (25) have a solution is just that $c(S_{\sigma,\tau}) \sim 1$ in $\mathcal{K}$. This is condition (II) of Theorem 14, while (24) is just condition (I) of that theorem. We thus have proved

THEOREM 16. *If a finite separable normal extension K of F has Galois groups G over F and H over $F(\mathcal{K})$, then the group of classes of crossed characters of G in $\mathcal{K}$ is isomorphic to the group of all those characters of H in $\mathcal{K}$ which satisfy conditions* (I) *and* (II) *of Theorem* 14.

In any event this shows that there can be but a finite number of classes of crossed characters. Suppose now that conditions (I) and (II) hold for all characters of H in $\mathcal{K}$. This means that there is a full complement of crossed characters, equal to the number of characters of H in $\mathcal{K}$. When H is Abelian and $\mathcal{K}$ has all requisite roots of unity, this means that the number of classes of crossed characters is just $[K:F(\mathcal{K})]$. But Theorem 3 showed that the number of classes of crossed characters is $[(R):(F)]$. Therefore $[(R):(F)] = [K:F(\mathcal{K})]$, which is enough to insure that K is a radical extension. This proves

COROLLARY 1. *A finite separable normal extension K/F is a radical extension if and only if the following conditions all hold:*

(i) *The group H of $K/F(\mathcal{K})$ is Abelian, with exponent e prime to the characteristic of $\mathfrak{F}$;*

(ii) *All e-th roots of unity are present in $\mathcal{K}$;*

(iii) *The Galois group satisfies the conditions* (I) *and* (II) *of Theorem* 14.

In other words, a radical extension can be identified by consulting simply properties of its group, of its coefficient extension, and of the portion of its group leaving invariant this coefficient extension.

CHAPTER III

ARITHMETIC OF SPECIAL RADICAL EXTENSIONS

11. **Finite coefficient fields.** In this chapter we aim to show that the general determination of all possible Galois groups for radical extensions, as given in Chapter II, can be used to get explicit formulas for these groups in a variety of special cases.

Let $\mathfrak{F}$ be a finite field of q elements, algebraically closed in F, while K is a finite separable radical extension of F. The coefficient field $\mathcal{K}$ of K then has q^f elements, where $f = [\mathcal{K}:\mathfrak{F}]$; it consists of 0 and of powers of a suitable single element ξ. $\mathcal{K}/\mathfrak{F}$ is cyclic, with a generating automorphism λ which maps any a in $\mathcal{K}$ on a^q. The generating equations for the radical extension $K = F(\mathcal{K}, r_1, \cdots, r_h)$ may now be written as

$$r_i^{m_i} = \xi^{g_i} x_i, \qquad i = 1, \cdots, h, \qquad g_i \text{ an integer.} \tag{1}$$

Application of the conditions of Theorem 12 shows that K/F is normal if and only if

$$q^f - 1 \equiv (q-1)g_i \equiv 0 \pmod{m_i}, \qquad i = 1, \cdots, h. \tag{2}$$

In the group $(R)/(F)$ of divisor classes associated with K we let $r^{\mathfrak{s}}$ denote the class belonging to the radical r. The Galois group of a normal K/F may be determined explicitly in terms of $(R)/(F)$, the constants q and f, and the parameters g_i of the generation (1), as follows:

THEOREM 17. *G is an extension of its Abelian normal subgroup $H = (R)/(F)$ by a cyclic group of order f. For a suitable representative U of the generator of this cyclic factor group, G is determined by the formulas*

$$U r^{\mathfrak{s}} U^{-1} = (r^{\mathfrak{s}})^q, \tag{3}$$

$$U^f = (r_1^{\mathfrak{s}})^{g_1} \cdots (r_h^{\mathfrak{s}})^{g_h}. \tag{4}$$

Proof. First, the subgroup $H = (R)/(F)$ is generated by automorphisms corresponding to the characters C_i with

$$C_i(r_i) = \xi^{t_i}, \qquad C_i(r_j) = 1, \qquad i \neq j. \tag{5}$$

Here $t_i = (q^f - 1)/m_i$ is an integer, by (2), so that ξ^{t_i} is a primitive m_i-th root of unity. In the second place, the automorphism $a \leftrightarrow a^q$ of $\mathcal{K}/\mathfrak{F}$ may be extended by Lemma 5 to an automorphism U for which

$$r_i^U = \xi^{s_i} r_i, \qquad s_i \equiv (q-1)g_i/m_i,$$

where s_i is an integer by (2). One then computes that U^f multiplies r_i by $\xi^{g_i t_i}$, so that the automorphism U of H corresponds in effect to the character $C = C_1^{g_1} \cdots C_h^{g_h}$, which does map r_i on $\xi^{g_i t_i}$. Replacement of the characters

by the isomorphic classes of divisors r^s gives the equation (4), while (3) is immediate.

This result is of interest because of the close parallel to the groups arising for p-adic fields. In the special case when the group $(R)/(F)$ of divisor classes is cyclic, the Galois group G described by (3) and (4) is exactly the Galois group of a normal extension of a p-adic field with a finite residue class field $\mathfrak{F}$ and with ramification order prime to p. The structure of these latter groups is given in Albert [1], Theorem 9.

12. **Universal groups over finite fields.** Instead of describing the possible Galois groups of a radical extension directly, as we have done, one may also try to describe them as the possible homomorphic images of a "universal" group which is the group of some universal (infinite) radical extension. In this description the universal group can even be replaced by a subgroup everywhere dense in the whole group, as was the case in Schilling's investigations of certain regular extensions of complete fields [14]. This construction will be carried out below for the case of a finite coefficient field.

Let F be a function field over its algebraically closed *finite* subfield $\mathfrak{F}$, of q elements, while D is a divisor group of F, of order prime to the characteristic, and with a basis $d_1, \cdots, d_h$ for $D/(F)$. Let d_i have order m_i, and choose x_i in F so that

$$d_i^{m_i} = (x_i), \qquad i = 1, \cdots, h. \tag{6}$$

Embed $\mathfrak{F}$ in its algebraically complete algebraic extension $\mathfrak{W}_\infty$, and let $W_\infty = F(\mathfrak{W}_\infty)$ be the corresponding coefficient extension of F. The group D can be interpreted as a group of divisors in W_∞, and Theorem 7 asserts that there is one and only one radical extension of W_∞ with divisor group D. This field K_∞ could also be described as the unique radical extension of F with divisor group and with coefficient field $\mathfrak{W}_\infty$; it may be generated, as $K_\infty = K(\mathfrak{W}_\infty, s_1, \cdots, s_h)$, by special radicals s_i with

$$s_1^{m_1} = x_1, \quad \cdots, \quad s_h^{m_h} = x_h. \tag{7}$$

This field K_∞ satisfies the conditions for normality over F (Theorem 12). Any radical extension of F with the same associated divisor group D_∞ as K_∞ is, by Theorem 7, a subfield of K_∞; indeed, K_∞ might be defined as the composite of all radical extensions of F having this D. This field K_∞, described in any one of these fashions, will be our *universal field, for given F and D*.

To determine the Galois group of K_∞, we first consider the group of its coefficient extension $\mathfrak{W}_\infty$. This involves the $n!$ integers (van Dantzig [5]), obtained by imposing on the ring of ordinary integers the topology in which a complete system of neighborhoods of zero is given by the principal ideals $(2!)$, $(3!), \cdots, (n!), \cdots$. The integers, completed with respect to this topology, yield the $n!$-integers; each such integer α is then a limit of a sequence of ordinary integers α_i, so that for any given integer f, α_i will have a constant remainder

modulo f for i sufficiently large. For any constant a in the field $\mathfrak{W}_\infty$ (with $a^{q^f} = a$ for some f), the sequence a^{e_i}, with $e_i = q^{\alpha_i}$, will then ultimately be a fixed element of $\mathfrak{W}_\infty$, which we write as

$$a^{q^\alpha} = \lim_{i\to\infty} a^{q^{\alpha_i}} \qquad (\alpha = \lim_{i\to\infty} \alpha_i).$$

The correspondence $a \to a^{q^\alpha}$ is an automorphism of $\mathfrak{W}_\infty$ which may be regarded as the symbolic power σ^α of the automorphism $a \to a^\sigma = a^q$. These automorphisms

$$a \to a^{\sigma^\alpha} = a^{q^\alpha}, \qquad a \in \mathfrak{W}_\infty, \qquad \alpha \text{ an } n!\text{-adic integer}, \tag{8}$$

constitute the whole Galois group of $\mathfrak{W}_\infty/\mathfrak{F}$, so that this group is an "ideal cyclic group". They are multiplied by the rule $\sigma^\alpha\sigma^\beta = \sigma^{\alpha+\beta}$, and the topology in the Galois group is exactly the topology given for the exponents α, as one may see by approximating $\mathfrak{W}_\infty$ by the sequence of finite subfields $\mathfrak{F}_n$ by degree $n!$ over $\mathfrak{F}$. The Galois group of K_∞ may now be described, as follows, in terms of the generating elements of (7).

THEOREM 18. *The extension K_∞/F has for each character C of its divisor class group $D/(F)$ an automorphism $u \to u^C$ described by the specifications that*

$$s_i^C = C[(s_i)]s_i, \qquad a^C = a \qquad (a \in \mathfrak{K}). \tag{9}$$

For each $n!$-adic integer α there is also an automorphism U^α of K_∞/F with

$$s_i^{U^\alpha} = s_i, \qquad a^{U^\alpha} = a^{q^\alpha} \qquad (a \in \mathfrak{K}). \tag{10}$$

The Galois group G_∞ of K_∞/F consists of all the products CU^α, multiplied by the table

$$U^\alpha U^\beta = U^{\alpha+\beta}, \qquad U^\alpha C U^{-\alpha} = C^{q^\alpha}, \tag{11}$$

where the C's are multiplied like characters. The topology in G_∞ is determined by taking for each integer n a neighborhood of the identity consisting of all $S = CU^\alpha$ with $C = I$ and $\alpha \equiv 0 \pmod{n!}$.

Proof. Since the field K_∞ may be regarded as the composite of the two subfields $F(\mathfrak{W}_\infty)$ and $F(s_1, \cdots, s_h)$, the formulas (9) and (10) are sufficient to completely determine these automorphisms C and U^α. Each automorphism σ^α of $\mathfrak{W}_\infty/\mathfrak{F}$ has a unique extension U^α to an automorphism of K_∞/F which will leave fixed the elements $s_1, \cdots, s_h$ which generate K_∞ over $F(\mathfrak{W}_\infty)$; this extension U^α has the properties of (10). By Theorem 13 the group of K_∞ over $F(\mathfrak{W}_\infty)$ is essentially the group H of all characters C, as in (9). Every automorphism does have the form CU^α, since the U^α exhaust the automorphisms of $F(\mathfrak{W}_\infty)/F$, and the formulas (11) follow from the corresponding formulas for finite extensions. Finally, the neighborhoods of the identity in G_∞ may be determined as the sets of automorphisms which leave fixed the respective finite subfields generated by $s_1, \cdots, s_h$ and the subfields $\mathfrak{F}_n$ of $\mathfrak{W}_\infty$. This gives the topology, as described in the theorem.

Consider the subgroup G_0 of G_∞ which is generated by the subgroup H of characters and the automorphism U. This subgroup G_0 then consists of the distinct automorphisms CU^k, for C any character of $D/(F)$, and k any integer. The multiplication table is simply

$$UCU^{-1} = C^q \qquad (C \text{ in } H), \tag{12}$$

so G_0 is a cyclic extension of the character group H of D.

THEOREM 19. *The Galois group G of any finite separable normal radical extension K with the divisor group D is a homomorphic image of the group G_0 of* (12).

Proof. Since $K \subset K_\infty$, by the universality of K_∞, the group G is certainly a homomorph of G_∞. Hence we need only show that every automorphism of G can be induced by an automorphism of the subgroup $G_0 \subset G_\infty$. This may be proved, either by appeal to the known structure of G in terms of characters (Chapter II) or by observing that G_0 is everywhere dense in the topology of G_∞. Since a subgroup is everywhere dense in an infinite Galois group if and only if the subgroup induces all automorphisms of every finite extension (Schilling [14], Lemma 5), the desired result follows. It means that G_0 is a "universal" Galois group for finite radical extensions with the group D.

It is possible to construct various other universal groups. For example, one may consider only those radical extensions in which the coefficient extension $\mathcal{K}$ is generated over $\mathfrak{F}$ by e-th roots of unity, where e is the exponent of D. For a fixed finite group D these fields are all contained in a universal field which has a *finite* universal group.

13. Restricted radical extensions. The question as to the variety of possible Galois groups for a radical extension may be formulated more sharply if one restricts attention to those extensions in which the attendant coefficient extension is as small as possible. If D is a given finite group of divisors, with exponent e, from a field F, then every normal radical extension in which the divisors of D become principal necessarily contains the field $\mathcal{K}_0 = \mathfrak{F}(\zeta)$ generated by a primitive e-th root of unity ζ. We say that the radical extension K is *restricted* if it is normal and separable and if its coefficient extension $\mathcal{K}$ is exactly this field $\mathcal{K}_0$. For given D, the Galois group of such a restricted extension K is obtained from two known components:

First, the group H, which is essentially the group of all linear characters C of $D/(F)$ in $\mathcal{K}_0$.

Second, the group Γ of $\mathcal{K}_0/\mathfrak{F}$. If $\mathfrak{k}$ is the intersection of $\mathfrak{F}$ and the subfield of $\mathcal{K}_0$ generated by ζ, Γ may also be described as the Galois group of $\mathfrak{k}(\zeta)/\mathfrak{k}$. $\mathfrak{k}$ itself is a finite field or a field of algebraic numbers, so Γ is known, as it is a subgroup of the appropriate cyclotomic group.

Our problem is the determination of those extensions G of H by Γ which can be realized as Galois groups of restricted radical extensions. This problem can be treated in two parts; first, what are the different classes of (similar) factor sets for such extensions; second, which of these factor sets satisfy the requisite

condition (II) of Theorem 14? The first part is purely group-theoretic, and the second is essentially a question in the splitting of a linear algebra determined by the factor set. In the case when $\mathfrak{F}$ has a finite characteristic, the Wedderburn theorem that there are no proper division algebras over a finite field insures that the factor sets are always similar to 1 in $\mathfrak{P}(\zeta)$ and hence in $\mathfrak{K}_0$, so that the second part of the problem is vacuous in this case.

We shall consider in particular conditions under which the Galois group is uniquely determined by the group D. When this is the case, the Galois group is that extension of H by Γ in which the factor set is identically 1. By Theorem 14, this particular group extension can always be realized; for example, as the group of the restricted extension $K = F(\mathfrak{K}_0, s_1, \cdots, s_h)$ generated by the special radicals s_i of (7).

14. Group construction when the exponent is an odd prime power.

THEOREM 20. *If the group $D/(F)$ of divisor classes is cyclic of order p^n, where p is an odd prime, then the Galois group G of a restricted radical extension with divisor group D is uniquely determined by the coefficient field $\mathfrak{F}$ and the order $e = p^n$ of D.*

Proof. Let Z be the multiplicative group generated by ζ, a primitive p^n-th root of unity. Every automorphism of Γ is then an automorphism of the group Z; since the latter automorphisms constitute the cyclic group of residues mod p^n which are prime to the odd prime p, the group Γ must be cyclic. Let a generator of Γ be the automorphism σ with

$$\zeta^\sigma = \zeta^k, \qquad k \text{ an integer.} \tag{13}$$

First consider the order of the automorphism σ of (13). By cyclotomic theory, its order must be a divisor of $\phi(p^n) = (p-1)p^{n-1}$. Suppose in particular that the integer k of (13) has $k \equiv 1 \pmod{p}$. Write

$$k = 1 + up^s, \qquad u \not\equiv 0 \pmod{p}, \qquad s \geqq 1. \tag{14}$$

Because p is odd (or, for $p = 2$ and $s \geqq 2$) the binomial theorem will give

$$k^p = 1 + vp^{s+1}, \qquad v \not\equiv 0 \pmod{p}.$$

Repeated applications of this, for an exponent $m \not\equiv 0 \pmod{p}$, prove

$$k^{p^i m} = 1 + wp^{s+i}, \qquad w \not\equiv 0 \pmod{p}. \tag{15}$$

The order of the automorphism σ of (13) is the least power $p^i m > 0$ for which $k^{p^i m} \equiv 1 \pmod{p^n}$. We thus conclude that *if* $k \equiv 1 \pmod{p^s}$ *and* $k \not\equiv 1 \pmod{p^{s+1}}$, *the order of the automorphism* σ *is* p^{n-s}.

The group Z of p^n-th roots of unity may be regarded as the group of characters of the cyclic group $D/(F)$ of divisor classes. By our general theory, the Galois group G of a restricted radical extension is thus a group extension of the cyclic group Z by the cyclic group Γ of order g. Since Γ is cyclic, G may be

represented as the set of all elements $\zeta^i U^j$, $0 \leqq i < p^n$, $0 \leqq j < g$, where U is a representative of the generating element σ of Γ, with a multiplication table

$$U\zeta^i U^{-1} = \zeta^{\sigma i}, \qquad U^g = c. \tag{16}$$

Here g is the order of Γ, and c is a constant of Z which satisfies the condition

$$c^\sigma = c. \tag{17}$$

By changing the choice of the representative U of σ, c may be changed to

$$c' = cb^{1+\sigma+\cdots+\sigma^{g-1}} = cNb, \tag{18}$$

where b may be any element in Z. We propose to show that G is uniquely determined as the extension with factor set 1; that is, with $c = 1$.

Suppose first that in (13) $k \not\equiv 1 \pmod p$, and that $c = \zeta^y$. The condition (17) then gives $\zeta^{y\sigma - y} = \zeta^{yk-y} = 1$, hence $y(k-1) \equiv 0 \pmod{p^n}$. Since $k - 1 \not\equiv 0 \pmod p$, the only solution for y is $y \equiv 0 \pmod{p^n}$. Thus $c = \zeta^y = 1$, as asserted.

Suppose next that $k \equiv 1 \pmod{p^s}$, but $k \not\equiv 1 \pmod{p^{s+1}}$. Then the solutions $c = \zeta^y$ of the condition (17) have $y \equiv 0 \pmod{p^{n-s}}$. On the other hand, in (18) the norm Nb of $b = \zeta^x$ is ζ with an exponent $x(k^g - 1)/(k-1)$. Since the order g of σ in this case is p^{n-s}, the exponent of ζ in this formula may by (14) and (15) be rewritten as $xwp^{s+n-s}/(up^s) = xp^{n-s}w/u \equiv 0 \pmod{p^{n-s}}$. $c = \zeta^y$ may thus be regarded as a norm $N\zeta^x$. Therefore in (18) the constant c' may be reduced to 1.

COROLLARY. *For a restricted radical extension with a divisor group D of odd prime power order p^n, the condition* (II) *of Theorem* 15 *for the realization of a Galois group is always satisfied.*

Proof. In this case the group $D/(F)$ is no longer cyclic, but the Galois group Γ of the coefficient field is still cyclic, just as above. The factor set for the Galois group G then can again be put in a form $U^g = S$, as in (16). The condition (II) for the realization of G then requires that each character $C(S)$ of S be a norm. But the value of this character $C(S)$ is a root of unity $c = \zeta^y$, which by the condition (I) must satisfy the analogue of (17). The computation above then shows that $C(S)$ is indeed a norm.

In this non-cyclic case, all potential groups can be realized, but these groups are no longer all identical, as in Theorem 20. We show this by an example. For an odd prime p first construct as coefficient field $\mathfrak{F}$ the field of characteristic ∞ [14] generated by all p-th roots of unity. If ζ is a primitive p^2-th root of unity, the degree of $\mathfrak{F}(\zeta)/\mathfrak{F}$ is then p. Let $F = \mathfrak{F}(x)$ be the field of rational functions in one indeterminate over $\mathfrak{F}$; in it one may construct a fractional divisor group D for which $D/(F)$ is Abelian of type (p^2, p). D has a basis of divisors d_1 and d_2 of orders p^2 and p, respectively. Its character group H is generated by characters C_1 and C_2 with

$$C_1(d_1) = \zeta, \quad C_1(d_2) = 1; \qquad C_2(d_1) = 1, \quad C_2(d_2) = \zeta^p.$$

[14] A similar example may be constructed for prime characteristic.

The potential Galois groups are then extensions of the group H of these characters by the cyclic group Γ generated by the automorphism σ of (13), where in this case $k \equiv 1 \pmod{p}$. As in (16), any such extension is given by a multiplication table $U^p = C$. One computes readily that $C = C_2$ is an allowable value for this constant ((17) holds), but that C_2 is not the norm of another character; hence the extension with this constant C_2 does not have factor set 1.

We observe that the technique used in this example can be readily expanded to give a computation for the *number* of different possible Galois groups for given D and $\mathfrak{F}$, in the case of a restricted radical extension.

One may show in other ways that the uniqueness assertion of Theorem 20 will not extend without restriction to the case when D is not of prime power exponent. For example, if $\mathfrak{F}$ is a finite field of 25 elements, while D is a cyclic group of order $9 \cdot 19 = 171$, one may show that $\mathfrak{F}(\zeta)/\mathfrak{F}$ has degree 9. In (18), the quantities Nb are all 1, but in (17) one may use for c any cube root of unity. In this case there are therefore three distinct extensions of Z by Γ, all three of which may be realized as Galois groups of restricted radical extensions.

15. **Group construction in the case 2^n.** The case of a cyclic divisor group of even prime power order must be given a separate treatment, in which the properties of quaternion algebras play an interesting role. It is desirable to treat separately the case when the base field has a prime characteristic.

THEOREM 21. *Let $D/(F)$ be a cyclic group of divisor classes of order 2^n, while $\mathfrak{F}$ is a coefficient field of characteristic ∞. Let $\mathfrak{L}$ be the intersection of $\mathfrak{F}$ with the field generated by the 2^n-th roots of unity. If $\mathfrak{L}$ is not a real field, only one Galois group G can be realized by a restricted radical extension with divisor group D, coefficient field $\mathfrak{F}$. If $\mathfrak{L}$ is real, and if $a^2 + b^2 = -1$ has no solution with elements a, b in $\mathfrak{F}$, then again only one group G can be realized. Finally, if $\mathfrak{L}$ is real, but $a^2 + b^2 = -1$ has a solution in $\mathfrak{F}$, then exactly two distinct groups G can be realized.*

Briefly speaking, this result shows that the Galois group in the prime power case is not always uniquely determined.

Proof. Let $\mathfrak{C}$ be the field generated (over the rational numbers) by the 2^n-th roots of unity. The whole Galois group of $\mathfrak{C}$ is generated by two automorphisms, α and β, with

$$\zeta^\alpha = \zeta^{-1}, \qquad \zeta^\beta = \zeta^5 \qquad (\zeta \text{ a primitive } 2^n\text{-th root}). \tag{19}$$

One has then

$$\alpha^2 = 1, \qquad \beta^{2^{n-2}} = 1, \qquad \alpha\beta = \beta\alpha. \tag{20}$$

If $n \geqq 3$, the whole group is Abelian of type $(2^{n-2}, 2)$; if $n = 2$, it is cyclic with generator α; if $n = 1$, it consists of the identity alone.

First we shall consider the possible group extensions of the cyclic group Z by the group Γ, since Γ is a subgroup of the whole group generated by α and β. We consider various subcases.

Case 1. Γ cyclic, generated by α. The factor set is then given as in (16), with $\sigma = \alpha$. The possible constants c, with $c^a = c$, are just $c = \pm 1$. The norm Nb of (18) is always 1, for by (19) $N\zeta = \zeta^{1+\alpha} = \zeta\zeta^{-1} = 1$. Hence in this case there are two different extensions, with $c = \pm 1$.

Case 2. Γ cyclic, generated by $\sigma = \beta^{2^k}$. By (19), the exponent β has the same effect as the exponent 5, and one may show that $5^{2^i} = 1 + u2^{i+2}$, where $u \not\equiv 0 \pmod 2$. It follows that the automorphism σ generating Γ has order 2^{n-2-k}. A straight-forward computation then proves that every invariant element c as in (17) has $c = \zeta^y$ with $y \equiv 0 \pmod{2^{n-2-k}}$, and that every such element is a norm. We conclude that in this case there is only one group extension.

Case 3. Γ cyclic, generated by $\sigma = \alpha\beta^{2^k}$. The order of Γ is then again 2^{n-2-k}, and the invariant elements are ζ^y with $y \equiv 0 \pmod{2^{n-1}}$. Each such is a norm $N_\sigma b$, so there is but one group extension.

These three cases are the only ones in which the group Γ is cyclic. The possible non-cyclic groups can all be treated together, as follows.

Case 4. Γ not cyclic, generated by α and $\sigma = \beta^{2^k}$. In such a case the group extension G of Z by Γ may be described by choosing representatives U_α and U_σ for the two generators α and σ of Γ; they will have a multiplication table

$$U_\alpha \zeta U_\alpha^{-1} = \zeta^\alpha, \qquad U_\sigma \zeta U_\sigma^{-1} = \zeta^\sigma, \tag{21}$$

$$U_\alpha^2 = c_\alpha, \qquad U_\sigma^{2^s} = c_\sigma, \qquad U_\sigma U_\alpha = a U_\alpha U_\sigma; \tag{22}$$

where $s = n - k - 2$ is the order of σ. The constants c_α, c_σ, a of (22) must satisfy certain associativity conditions (see Zassenhaus [21], p. 97). One of these conditions is that $c_\alpha^\alpha = c_\alpha$; it implies as in Case 1 above that $c_\alpha = \pm 1$. The value of c_α cannot be altered by the choice of a new representative for U_α. Another condition is $c_\sigma^\sigma = c_\sigma$, as in (17); by the technique of Case 2, this enables us to make $c_\sigma = 1$ by choosing a new representative for U_σ. After this reduction has been carried out, one of the associativity conditions for the third constant a (obtained from the equation $U_\sigma^{2^s} U_\alpha = U_\alpha$) is

$$1 = N_\sigma a = a^{1+\sigma+\sigma^2+\cdots+\sigma^{s-1}}.$$

If $a = \zeta^z$, one may compute that a satisfies this condition if and only if $z \equiv 0 \pmod{2^{2+k}}$. But the replacement of the representative U_α by $U'_\alpha = bU_\alpha$ changes the constant a to $ab^{\sigma-1}$. By a suitable choice of the constant b, this result will be $ab^{\sigma-1} = 1$. After these reductions, the multiplication table takes on one of two forms

$$U_\alpha^2 = \pm 1, \qquad U_\sigma^{2^s} = 1, \qquad U_\sigma U_\alpha = U_\alpha U_\sigma. \tag{23}$$

Again in this case there are exactly two possible group extensions.

The subdivision into cases may now be reformulated. There are two possible group extensions in all those instances when the group Γ contains the automorphism α; in the remaining cases there is but one extension. Now α is

that automorphism which maps each root of unity into its complex conjugate, so the subfield of all elements invariant under α is just the maximal real subfield of the cyclotomic field. Since Γ is the group of $\mathfrak{L}(\zeta)$ over $\mathfrak{L}$, where $\mathfrak{L}$ is the intersection of $\mathfrak{F}$ and the cyclotomic subfield, it follows that there will be two distinct group extensions if and only if $\mathfrak{L}$ is a real field. (Observe that if $\mathfrak{L}$ is real, all of its conjugates are real, so that this statement does not involve the possible ambiguity in the determination of $\mathfrak{L}$.)

In those cases when two group extensions are possible it remains to determine when the corresponding groups can both be realized. By the condition (II) of Theorem 15 this is just a question about algebras: when is the crossed product algebra $A = (\mathfrak{F}(\zeta), \Gamma, C(S_{\sigma,\tau}))$ a total matric algebra? Here $S_{\sigma,\tau}$ is the factor set for the group extension, $C(S_{\sigma,\tau})$ is the corresponding set of characters. In the Case 4 treated above, this factor set is determined by the constants $c_\alpha = \pm 1$, $c_\sigma = 1$, $a = 1$ of (22) and (23). Because of the simple form of these constants, the algebra may be reduced. The whole group Γ is the direct product of the cyclic subgroup of order 2 generated by α and the cyclic subgroup Δ generated by $\sigma = \beta^{2^k}$. One may show that the factor set determined by (22) and (23) is "symmetric" relative to the subgroup Δ; a general theorem on crossed product algebras (see, e.g., Schilling-Mac Lane [12]) then asserts that the whole algebra is similar to another algebra with the *same* factor set, restricted to the group Γ/Δ. Now Γ/Δ is just the cyclic group of order 2 generated by α. Hence the algebra under consideration is just the cyclic algebra

$$A \sim (\mathfrak{F}(i), \alpha, \pm 1),$$

with the multiplication table

$$U_\alpha i U_\alpha^{-1} = -i, \qquad U_\alpha^2 = \pm 1,$$

where i is a primitive 4-th root of unity. If the constant U_α^2 here is $+1$, this is a total matric algebra; if it is -1, it is just the ordinary algebra of quaternions over the field $\mathfrak{F}$. The latter algebra will be a total matric algebra if and only if its multiplication constant -1 is a norm from $\mathfrak{F}(i)$:

$$-1 = N(a + bi) = a^2 + b^2 \qquad (a, b \in \mathfrak{F}).$$

Therefore the group in question can be realized by a radical extension if and only if the equation $a^2 + b^2 = -1$ has a solution in $\mathfrak{F}$. In the remaining Case 1, the same condition may be found, without the necessity of a preliminary reduction of the algebra. We thus have all the results stated in Theorem 21.

Remark. In considering the condition that the algebra A be a total matric algebra over the coefficient field $\mathfrak{F}$, one might be tempted to try to reduce this question to one about algebras over an algebraic number field, for in this case the invariants of an algebra are completely known. If $\mathfrak{F}_0$ is the field of all algebraic numbers contained in the given field $\mathfrak{F}$, the algebra $A = (\mathfrak{K}, \Gamma, C(S_{\sigma,\tau}))$

is a scalar extension of the algebra $A_0 = (\mathfrak{F}_0(\zeta), \Gamma, C(S_{\sigma,\tau}))$, and the base field $\mathfrak{F}_0$ of A_0 is algebraically closed in the extended base field $\mathfrak{F}$ of A. However, it would be false to suppose that A is a total matric algebra if and only if A_0 is one.

This may be supported by an example. Let $\mathfrak{P}$ be the field of rational numbers, and, as in Witt ([19], p. 10), let $\mathfrak{F} = \mathfrak{P}(x, y)$ with $y^2 = -1 - x^2$ be a function field of one variable over $\mathfrak{P}$. One may show that $\mathfrak{P}$ is algebraically closed in $\mathfrak{F}$. However, the quaternion algebra considered above does not split over $\mathfrak{P}$, and does split over $\mathfrak{F}$, for $-1 = x^2 + y^2$ is a sum of two squares by the very definition of $\mathfrak{F}$.

To complete our investigation of the prime power case we state a result parallel to the previous theorem, but for fields of finite characteristic. In the statement, ζ is again a primitive 2^n-th root of unity over the field $\mathfrak{F}$, while $\mathfrak{k}$ is the intersection of $\mathfrak{F}$ and the field generated by ζ alone.

THEOREM 22. *Let $D/(F)$ be a cyclic group of divisor classes of order 2^n, while $\mathfrak{F}$ is a field of finite characteristic p prime to 2. The Galois group G of a restricted radical extension with divisor group D is uniquely determined by $\mathfrak{F}$ and 2^n, except in the case when*

$$n > 1, \qquad \mathfrak{k} = \mathfrak{P}, \qquad p \equiv -1 \pmod{2^n},$$

where $\mathfrak{P}$ is the prime field of characteristic p. In this exceptional case there are exactly two groups possible.

The proof uses the same methods as the previous theorems, so will be omitted. Incidentally, the exceptional case can also be described as that case (with $n > 1$) in which $x^2 = -1$ has no solution in $\mathfrak{F}$, while the field $\mathfrak{F}(\zeta)$ generated by a primitive 2^n-th root of unity contains a primitive 2^{n+1}-th root of unity.

16. **Function fields of one variable.** From the classical point of view, the most important special case of our theory is that in which F is a function field of *one* variable over a coefficient field $\mathfrak{F}$. In this case one gives special attention to the ordinary or *arithmetic* divisors $A = P_1^{e_1} \cdots P_n^{e_n}$; as described in the introduction, these are obtained from the prime divisors P_i (or from the equivalent valuations) of F. Those divisors A which become principal in any finite extension of F necessarily have finite order (i.e., some A^m is the divisor of a function), as was proved in the introduction. But the divisors A of finite order constitute a group $\mathfrak{D}_\infty$ of *arithmetic* divisors which may be regarded as a subgroup of the group D_∞ of fractional divisors. One would then be led to consider those radical extensions for which the associated group D of divisors consists exclusively of such arithmetic divisors. These extensions can be characterized as follows.

THEOREM 23. *A separable radical extension K of a function field F of one variable is unramified if and only if its associated group D of fractional divisors consists entirely of arithmetic divisors.*

Proof. Recall that a prime divisor P of F is said to be unramified in an extension K if P can be decomposed into a product of *distinct* prime divisors of K. The whole extension K is unramified if every prime divisor of F is unramified.

First, suppose that K is a radical extension in which a non-arithmetic fractional divisor d becomes principal. If d has order m, with $d^m = (x)$, the arithmetic divisor $(x) = P_1^{e_1} \cdots P_n^{e_n}$ then involves at least one prime divisor P_i to an exponent e_i not divisible by m. Let V_i be the valuation of F corresponding to the prime divisor P_i, so that for each function y of the field, $V_i y$ is the exact power of P_i dividing y. The radical r belonging to the given divisor d has $r^m = bx$, for b a constant. In any extension V_i' of the valuation to K, $V_i' r = V_i(bx)/m = e_i/m$. Since this is a proper fraction, the valuation V_i and hence the corresponding prime ideal P_i is ramified.

Conversely, suppose that K is a radical extension belonging to a group D of arithmetic divisors. It is well known that the decomposition of divisors of F can be found in terms of the corresponding decompositions of ideals in suitable integrally closed rings. Specifically, let z be any function of F transcendental over $\mathfrak{F}$, and let $\mathfrak{O}_z$ be the ring of all functions of F integral over $\mathfrak{F}[z]$, while $\mathfrak{O}_{1/z}$ is the ring of all functions integral over $\mathfrak{F}[1/z]$. Then[15] each prime divisor P of F corresponds either to a prime ideal $\mathfrak{p}$ of $\mathfrak{O}_z$ or to a prime ideal $\mathfrak{p}$ of $\mathfrak{O}_{1/z}$ which is a factor of $1/z$. Furthermore, P is ramified in an extension K if and only if the corresponding prime ideal $\mathfrak{p}$ is ramified in K (that is, has a square of some prime ideal as a factor). However, the ramified prime ideals from a ring $\mathfrak{O}$ are simply the prime ideal factors of the discriminant of the extension $\mathfrak{O}^*$ of $\mathfrak{O}$ ($\mathfrak{O}^*$ is the integral closure of $\mathfrak{O}$ in K). This discriminant is in turn a factor of the discriminant of any integral element of K generating K over $\mathfrak{O}$.

The radical extension K under consideration may now be proved to be unramified by examination of the discriminants of suitable generating elements. For example, consider a separable coefficient extension $K = F(\mathfrak{K})$. This can be generated by an element of which is integral with respect to either ring $\mathfrak{O}_z$ or $\mathfrak{O}_{1/z}$. The discriminant for this element is a constant of $\mathfrak{F}$; hence no prime ideals are ramified in this case. On the other hand let $K = F(r)$ be generated[16] by a radical r with an arithmetic divisor $(r) = A$. If $r^m = x$, for x in F, the discriminant of this generating equation, relative to $\mathfrak{O}_x$, will involve only prime ideal factors of x. On the other hand, the same extension $F(r)$ may be obtained by adjoining a different radical $r' = r/u$, where u is a function of F with a divisor $(u) = AB$, B relatively prime to A. The discriminant for this second defining equation will no longer involve the prime ideal factors of x. All told,

[15] See, for example, the discussion in F. K. Schmidt [15], pp. 6-7, and p. 10. His statements, which envisage the case of a finite coefficient field $\mathfrak{F}$, are also valid for an arbitrary coefficient field. In case $\mathfrak{F}$ is imperfect, the requisite Noether ideal theory in $\mathfrak{O}$ may be conveniently established by means of valuation theory, as in F. K. Schmidt [16], §3, 3.

[16] The fact that such an extension is unramified can also be proved by arguments with valuations, as in Deuring [6], Theorem 9 (for the case when the m-th roots of unity all lie in F).

one can conclude (with similar arguments for $\mathfrak{O}_{1/z}$) that no prime ideals are ramified.

The general radical extension K with an *arithmetic* divisor group is obtained from F by a succession of adjunctions of the two types considered above; hence a combination of the above arguments proves the theorem.

COROLLARY. *Let K be a normal extension of F, with the same coefficient field as F, in which a group $\mathfrak{D}$ of arithmetic divisors of F become principal. Assume that F contains all e-th roots of unity, where e is the exponent of $\mathfrak{D}$ (and is prime to the characteristic of F). Then every divisor of $\mathfrak{D}$ will become principal in an unramified Abelian subfield of K which has over F a Galois group isomorphic to $\mathfrak{D}/(F)$.*

Proof. The subfield in question is the field generated by all radicals of K with (r) in $\mathfrak{D}$. The subfield is unramified by the theorem above, while its Galois group is given by Theorem 13.

Remark. If K is an unramified normal extension of F with the same coefficient field as F, and if every root of unity appearing in the characters of the Galois group of K over F lies in the field F, then the maximal Abelian subfield of K can be described as the maximal subfield of K which is a radical extension of F.

17. **Examples of divisor groups in the arithmetic case.** To show that our theory of the Galois groups of radical extensions applies in full form to the case of extensions with an associated group of arithmetic divisors, we shall now show that one may realize groups of arithmetic divisors of arbitrary complexity in a suitably chosen function field.

THEOREM 24. *If $\mathfrak{F}$ is any coefficient field, $\mathfrak{D}^*$ any finite Abelian group, with order prime to the characteristic of $\mathfrak{F}$, then there exists a function field F of one variable over $\mathfrak{F}$ which contains a group $\mathfrak{D}$ of arithmetic divisors such that the corresponding group $\mathfrak{D}/(F)$ of divisor classes is isomorphic to the given group $\mathfrak{D}^*$, provided only that $\mathfrak{F}$ contains a sufficiently large (finite) number of elements.*

Proof. Since the given Abelian group $\mathfrak{D}^*$ can always be embedded in a direct product of t cyclic groups of order n, where n is prime to the characteristic, it will suffice to realize the latter group as a group of divisor classes. We shall assume that the coefficient field $\mathfrak{F}$ has at least $n + t$ distinct elements $a_1, a_2, \cdots, a_{n+t}$.

First construct a field $k = \mathfrak{F}(x)$ over $\mathfrak{F}$, and over this a function field $F = k(y) = \mathfrak{F}(x, y)$, where $y^n = (x - a_1)(x - a_2) \cdots (x - a_s)$, while s is an integer between $t + 1$ and $t + n$, so chosen that $(s, n) = 1$. In k let $(x - a_i) = p_i/p$, where $i = 1, \cdots, s$. In the valuation V_i corresponding to the prime divisor p_i, one has $V_i y = (1/n)V_i(x - a_i)$; hence p_i has the decomposition $p_i = P_i^n$ in F. Since s is prime to n, one also has $p_\infty = P_\infty^n$. We propose then to consider the divisors $A_i = P_i/P_\infty$ of order zero in F. These divisors become principal in the extended field $K = F((x - a_1)^{1/n}, \cdots, (x - a_s)^{1/n})$.

We shall prove that each divisor A_i has order n and that they generate together a group of divisor classes which is the direct product of $s-1$ cyclic groups of order n. For this purpose consider the extensions $k' = k(\zeta)$, $F' = F(\zeta)$, $K' = K(\zeta)$, where ζ is a primitive n-th root of unity. We remark that the A_i's become divisors in F' of exactly the same order as in F, according to Corollary 2 of Theorem 3. Let ω be the group of all these elements of k' which have an n-th root in the extension K'. Since the n-th roots of unity all lie in k', the ordinary Kummer theory (Witt [20]) asserts that the degree of K' over k' is $[\omega : k'^{*n}]$. Since ω contains each $(x - a_i)$,

$$[K' : k'] \geqq [k'^{*n}\{(x-a_1), \cdots, (x-a_s)\} : k'^{*n}].$$

By the ordinary decomposition theorem for the polynomials of $\mathfrak{F}'[x]$, the index on the right is n^s. On the other hand, the generation of K' by radicals shows that $[K' : k'] \leqq n^s$. Therefore $[K' : k'] = n^s$. The same argument shows that $[F' : k'] = n$; therefore $[K' : F'] = n^{s-1}$.

Over F', the radical extension K' can be generated by $s-1$ of the radicals, say by $(x-a_1)^{1/n}, \cdots, (x-a_{s-1})^{1/n}$. Furthermore the extension K' involves no coefficient extension over k' or over F', for, over any larger coefficient field $\mathfrak{W} \supset \mathfrak{F}$ one could again apply the above argument about degrees to show that $[K'(\mathfrak{W}) : k'(\mathfrak{W})] = n^s$; this implies that no proper extension $\mathfrak{W}$ could be contained in K'.

The group D of (arithmetic) divisors which become principal in K' is just the group generated by the divisors A_i of $(x-a_1)^{1/n}$, $i = 1, \cdots, s-1$, according to the Corollary of Theorem 5. By Theorem 5, this group D has an index $[D : (F')] = [K' : F'] = n^{s-1}$. In the original field F, the group of divisor classes generated by $P_1/P, \cdots, P_{s-1}/P$ therefore has the required structure.

This theorem shows that the consideration of non-Abelian extensions is essential. For suppose that the coefficient field $\mathfrak{F}$ of the theorem does not contain all n-th roots of unity. Since there nevertheless may be divisor classes of order n, a normal extension of F in which such a divisor class becomes principal cannot be Abelian, by the first Corollary of Theorem 15. Any such normal extension will necessarily involve a coefficient extension.

The case when the coefficient extension $\mathfrak{F}$ is finite is especially interesting, in view of the class field theory. For any given integer n, there will exist a prime $p \geqq n + t$ such that $p \not\equiv 1 \pmod{n}$. The finite field $\mathfrak{F}$ of p elements will then be one to which the construction of the theorem applies. By the existence theorem of the class field theory (Witt [20]), there will exist unramified Abelian extensions of F with a degree divisible by n. In none of these extensions will the divisor classes of order n in the field become principal.

HARVARD UNIVERSITY AND THE UNIVERSITY OF CHICAGO.

BIBLIOGRAPHY

1. A. A. ALBERT, *On p-adic fields and rational division algebras*, Annals of Mathematics, vol. 41 (1940), pp. 674–693.

2. E. Artin, *Idealklassen in Oberkörpern und allgemeines Reziprozitätsgesetz*, Abhandlungen aus dem Mathematischen Seminar, Hamburg, vol. 7(1930), pp. 46–51.
3. R. Baer, *Abelian fields and duality of Abelian groups*, American Journal of Mathematics, vol. 59(1937), pp. 869–888.
4. A. H. Clifford and S. Mac Lane, *Factor sets of a group in its abstract unit group*, Transactions of the American Mathematical Society, vol. 50(1941), pp. 385–406.
5. D. van Dantzig, *Nombres universels ou p-adiques avec une introduction sur l'algèbre topologique*, Annales de l'Ecole normale, (3), vol. 53(1936), pp. 257–307.
6. Max Deuring, *Zur arithmetischen Theorie der algebraischen Funktionen*, Mathematische Annalen, vol. 106(1932), pp. 77–102.
7. Ph. Furtwängler, *Beweis des Hauptidealsatzes für die Klassenkörper algebraischer Zahlkörper*, Abhandlungen aus dem Mathematischen Seminar, Hamburg, vol. 7(1930), pp. 14–36.
8. H. Hasse, *Theorie der relativ-zyklischen algebraischen Funktionenkörper, insbesondere bei endlichem Konstantenkörper*, Journal für Mathematik, vol. 172(1935), pp. 37–54.
9. S. Mac Lane, *A lattice formulation for transcendence degrees and p-bases*, this Journal, vol. 4(1938), pp. 455–468.
10. S. Mac Lane, *Modular fields*. I, *Separating transcendence bases*, this Journal, vol. 5(1939), pp. 372–393.
11. S. Mac Lane and O. F. G. Schilling, *Normal algebraic number fields*, Transactions of the American Mathematical Society, vol. 50(1941), pp. 295–384.
12. S. Mac Lane and O. F. G. Schilling, *A formula for the direct product of cross product algebras*, Bulletin of the American Mathematical Society, vol. 48(1942), pp. 108–114.
13. E. Noether, *Der Hauptgeschlechtssatz für relativ-galoissche Zahlkörper*, Mathematische Annalen, vol. 108(1933), pp. 411–419.
14. O. F. G. Schilling, *Regular normal extensions over complete fields*, Transactions of the American Mathematical Society, vol. 47(1940), pp. 440–454.
15. F. K. Schmidt, *Analytische Zahlentheorie in Körpern der Charakteristik p*, Mathematische Zeitschrift, vol. 33(1931), pp. 1–32.
16. F. K. Schmidt, *Zur arithmetischen Theorie der algebraischen Funktionen*, I, Mathematische Zeitschrift, vol. 41(1936), pp. 415–438.
17. F. K. Schmidt, *Die Theorie der Klassenkörper über einem Körper algebraischen Funktionen in einer Unbestimmten und mit endlichem Koeffizientenbereich*, Sitzungsberichte der Physikalisch-Medezinische Sozietät, Erlangen, vol. 62(1930), pp. 267–284.
18. A. Speiser, *Zahlentheoretische Sätze aus der Gruppentheorie*, Mathematische Zeitschrift, vol. 5(1919), pp. 1–6.
19. E. Witt, *Zerlegung reeller algebraischer Funktionen in Quadrate. Schiefkörper über reellem Funktionenkörper*, Journal für Mathematik, vol. 171(1934), pp. 4–11.
20. E. Witt, *Der Existenzsatz für abelsche Funktionenkörper*, Journal für Mathematik, vol. 173(1935), pp. 43–51.
21. H. Zassenhaus, *Lehrbuch der Gruppentheorie*, Hamburger Mathematische Einzelschriften, vol. 21(1937).

ABSOLUTE NÖRLUND SUMMABILITY

By Leonard McFadden

1. **Introduction.** The method of summability considered here was first introduced by Woronoi [11],[1] but is more closely identified with the name of Nörlund [8]. Accordingly, we will use the term "Nörlund transformations" to indicate the transformations in question.

Let $\{p_n\}$ be a sequence of constants, real or complex valued, and let $\{P_n\}$ denote the sequence of partial sums. We will call $\{t_n\}$ the Nörlund transform of an arbitrary sequence $\{U_n\}$, where

$$t_n = \sum_{v=0}^{n} \frac{p_{n-v} U_v}{P_n}, \tag{1.01}$$

it being assumed of course that $P_n \neq 0$. The sequence $\{U_n\}$ is said to be summable by the Nörlund mean N_p defined by $\{p_n\}$, or summable N_p, if $\lim_{n\to\infty} t_n$ exists.

The conditions for regularity of such a transformation are

$$\lim_{n\to\infty} \frac{p_n}{P_n} = 0, \tag{1.02}$$

$$\sum_{k=0}^{n} |p_k| \leqq C |P_n|, \tag{1.03}$$

where C is a finite positive constant. It is easily seen that (1.02) is equivalent to

$$\lim_{n\to\infty} \frac{P_{n-1}}{P_n} = 1. \tag{1.04}$$

We might also note that, if p_n is real and non-negative, condition (1.03) is satisfied automatically and, if in addition p_n is non-increasing, condition (1.02) is also satisfied.

The sequence $\{U_n\}$ will be said to be absolutely summable by the Nörlund mean defined by the sequence $\{p_n\}$, or summable $|N_p|$, provided that

$$\sum_{n=1}^{\infty} |t_n - t_{n-1}| \leqq C < \infty. \tag{1.05}$$

If we let N_p and $|N_p|$ denote, respectively, the class of sequences summable N_p and $|N_p|$, we have the following

(1.06) Theorem. $|N_p| \subset N_p$; $N_p \not\subset |N_p|$.

Received November 3, 1941.

[1] Numbers in brackets refer to the bibliography at the end of the paper.

Proof. To prove the first relation, we observe that, since $\{U_n\}$ is summable $|N_p|$, then $\sum_{n=0}^{\infty} |t_n - t_{n-1}| = C < \infty$. Therefore, for given $\epsilon > 0$ there exists a sufficiently large positive integer N such that for any $q > 0$ we have, when $n > N$, $\sum_{m=n}^{n+q} |t_m - t_{m-1}| < \epsilon$. It follows that $|t_{n+q} - t_n| = |\sum_{m=n+1}^{n+q} (t_m - t_{m-1})| \leqq \sum_{m=n}^{n+q} |t_m - t_{m-1}| < \epsilon$. Hence, t_n converges to some limit t. This is, however, the condition that $\{U_n\}$ be summable N_p.

To prove the second relation we exhibit the following example. Let $p_n = 1$, $n = 0, 1, \cdots$, and let $U_n = (-1)^n$, $n = 0, 1, \cdots$. Then

$$t_k = \begin{cases} \dfrac{1}{k+1}, & \text{for } k \text{ even}, \\ 0, & \text{for } k \text{ odd}. \end{cases}$$

Clearly t_k converges to zero, but

$$|t_k - t_{k-1}| = \begin{cases} \dfrac{1}{k+1}, & \text{for } k \text{ even}, \\ \dfrac{1}{k}, & \text{for } k \text{ odd}. \end{cases}$$

Hence, $\sum_{k=1}^{\infty} |t_k - t_{k-1}|$ diverges and our theorem is proved.

Since there is no loss of generality in considering $p_0 = 1$, we shall do so throughout and, unless otherwise stated, it will be assumed that the transformations considered are regular.

2. **Inclusiveness relations.** Kogbetliantz [6] proved that a series absolutely summable by the Cesàro mean of order α is also absolutely summable by the Cesàro mean of order $\beta > \alpha$. It is well known that the Cesàro transformation is a particular case of the Nörlund transformation. Thus we are led to the problem of determining what conditions must be imposed on the sequences $\{p_n\}$ and $\{q_n\}$ in order that a sequence summable $|N_q|$ will also be summable $|N_p|$.

We will first state the following result due to Florence Mears [7]. Given the matrix $||a_{nk}||$, let $U_n = \sum_{k=0}^{n} u_k$, $U'_n = \sum_{k=0}^{\infty} a_{nk} U_k$ and $u'_n = U'_n - U'_{n-1}$.

(2.01) THEOREM. *The necessary and sufficient conditions that* $\sum_{n=0}^{\infty} |u'_n|$ *converge whenever* $\sum_{n=0}^{\infty} |u_n|$ *converges are*

(1) $\sum_{k=0}^{\infty} a_{nk}$ *converges for all* n;

(2) $\sum_{n=0}^{\infty} \left| \sum_{v=k}^{n} (a_{nv} - a_{n-1v}) \right| \leqq C < \infty$ *for all* k.

Now consider

$$t_n = \sum_{v=0}^{n} \frac{p_{n-v} U_v}{P_n} \tag{2.02}$$

and

$$\tau_n = \sum_{v=0}^{n} \frac{q_{n-v} U_v}{Q_n}; \tag{2.03}$$

then

$$t_n = \sum_{v=0}^{n} \frac{R_{n-v} Q_v \tau_v}{P_n}, \tag{2.04}$$

where R_k is obtained from the following system of equations obtained by equating the coefficients of U_k in the numerators of the right hand sides of (2.02) and (2.04):

$$\begin{aligned} p_0 &= q_0 R_0, \\ p_1 &= q_1 R_0 + q_0 R_1, \\ &\cdots\cdots\cdots\cdots\cdots\cdots, \\ p_k &= q_k R_0 + \cdots + q_0 R_k. \end{aligned} \tag{2.05}$$

Adding, we get

$$P_k = Q_k R_0 + \cdots + Q_0 R_k. \tag{2.06}$$

If we write

$$p(x) = p_0 + p_1 x + \cdots + p_n x^n + \cdots, \tag{2.07}$$

$$q(x) = q_0 + q_1 x + \cdots + q_n x^n + \cdots, \tag{2.08}$$

$$R(x) = R_0 + R_1 x + \cdots + R_n x^n + \cdots, \tag{2.09}$$

it is easily seen that

$$\frac{p(x)}{q(x)} = R(x). \tag{2.10}$$

Wherever they appear, it will be understood that $p_{-1} = P_{-1} = R_{-1} = 0$.

(2.11) THEOREM. *A necessary and sufficient condition that* $|N_q| \subset |N_p|$ *is*

$$\sum_{n=0}^{\infty} \left| \sum_{v=k}^{n} \left(\frac{R_{n-v} Q_v}{P_n} - \frac{R_{n-v-1} Q_v}{P_{n-1}} \right) \right| \leqq C < \infty \qquad \textit{for all } k.$$

Proof. The matrix $\| R_{n-v} Q_v / P_n \|$ satisfies the conditions of Theorem (2.01), due to (2.06) and the hypothesis.

(2.12) THEOREM. *Sufficient conditions that* $|N_q| \subset |N_p|$ *are*

$$(1)\quad \sum_{v=k}^{n}\left(\frac{R_{n-v}Q_v}{P_n}-\frac{R_{n-v-1}Q_v}{P_{n-1}}\right)\geqq 0, \qquad k=0,1,\cdots,n;$$

$$(2)\quad \sum_{v=0}^{n}\left|\frac{R_{n-v}Q_v}{P_n}\right|\leqq C<\infty \qquad \text{for all } n.$$

Proof. It suffices to show that the condition of Theorem (2.11) is satisfied.

$$\sum_{n=0}^{N}\left|\sum_{v=k}^{n}\left(\frac{R_{n-v}Q_v}{P_n}-\frac{R_{n-v-1}Q_v}{P_{n-1}}\right)\right|=\sum_{n=0}^{N}\sum_{v=k}^{n}\left(\frac{R_{n-v}Q_v}{P_n}-\frac{R_{n-v-1}Q_v}{P_{n-1}}\right),$$

by (2.12), (1),

$$=\sum_{v=k}^{N}\frac{R_{N-v}Q_v}{P_n}\leqq\sum_{v=0}^{N}\left|\frac{R_{N-v}Q_v}{P_n}\right|\leqq C<\infty$$

for all N, by (2.12), (2). Therefore

$$\sum_{n=0}^{\infty}\left|\sum_{v=k}^{n}\left(\frac{R_{n-v}Q_v}{P_n}-\frac{R_{n-v-1}Q_v}{P_{n-1}}\right)\right|\leqq C<\infty \qquad \text{for all } k.$$

(2.13) THEOREM. *If* (1) $\{P_n\}$ *and* $\{Q_n\}$ *are both non-negative sequences,* (2) $\sum_{v=0}^{n}(|R_{n-v}|\,Q_v/P_n)\leqq C<\infty$ *independently of* n, *and* (3) *there exists a positive integer* N *such that* $(R_{m-1}/P_{n-1})-(R_m/P_n)\geqq 0$ *whenever* $n\geqq m\geqq N$, *then* $|N_q|\subset|N_p|$.

Proof. Again we show that the condition of Theorem (2.11) is satisfied. First,

$$(2.14)\quad \begin{aligned}&\sum_{n=0}^{\infty}\left|\sum_{v=k}^{n}\left(\frac{R_{n-v}Q_v}{P_n}-\frac{R_{n-v-1}Q_v}{P_{n-1}}\right)\right|\\&=\sum_{n=k}^{k+N}\left|\sum_{v=k}^{n}\left(\frac{R_{n-v}Q_v}{P_n}-\frac{R_{n-v-1}Q_v}{P_{n-1}}\right)\right|+\sum_{n=k+N+1}^{\infty}\left|\sum_{v=k}^{n}\left(\frac{R_{n-v}Q_v}{P_n}-\frac{R_{n-v-1}Q_v}{P_{n-1}}\right)\right|\end{aligned}$$

and

$$(2.15)\quad \begin{aligned}&\sum_{v=k}^{n}\left(\frac{R_{n-v}Q_v}{P_n}-\frac{R_{n-v-1}Q_v}{P_{n-1}}\right)\\&=\sum_{v=0}^{n}\frac{R_{n-v}Q_v}{P_n}-\sum_{v=0}^{k-1}\frac{R_{n-v}Q_v}{P_n}-\sum_{v=0}^{n-1}\frac{R_{n-v-1}Q_v}{P_{n-1}}+\sum_{v=0}^{k-1}\frac{R_{n-v-1}Q_v}{P_{n-1}}\\&=1-\sum_{v=0}^{k-1}\frac{R_{n-v}Q_v}{P_n}-1+\sum_{v=0}^{k-1}\frac{R_{n-v-1}Q_v}{P_{n-1}}, \qquad \text{by (2.06)},\\&=\sum_{v=0}^{k-1}Q_v\left(\frac{R_{n-v-1}}{P_{n-1}}-\frac{R_{n-v}}{P_n}\right).\end{aligned}$$

Now

$$\left|\sum_{v=0}^{k-1}\frac{Q_v R_{n-v-1}}{P_{n-1}}\right| \leqq \sum_{v=0}^{n-1}\frac{Q_v\,|\,R_{n-v-1}\,|}{P_{n-1}} \leqq C,$$

$$\left|\sum_{v=0}^{k-1}\frac{Q_v R_{n-v}}{P_n}\right| \leqq \sum_{v=0}^{n}\frac{Q_v\,|\,R_{n-v}\,|}{P_n} \leqq C, \tag{2.16}$$

by (2.13), (1), and (2.13), (2). Therefore,

$$\begin{aligned}\sum_{n=k}^{k+N}\left|\sum_{v=k}^{n}\left(\frac{R_{n-v}Q_v}{P_n}-\frac{R_{n-v-1}Q_v}{P_{n-1}}\right)\right| &= \sum_{n=k}^{k+N}\left|\sum_{v=0}^{k-1}Q_v\left(\frac{R_{n-v-1}}{P_{n-1}}-\frac{R_{n-v}}{P_n}\right)\right|, && \text{by (2.15),}\\ &\leqq (N+1)\,2C<\infty && \text{for all } k, \text{ by (2.16).}\end{aligned} \tag{2.17}$$

When $n \geqq k+N+1$, it is clear that $n-v \geqq N+1+k-v$, and so, by (2.13), (3), $(R_{n-v-1}/P_{n-1}) - (R_{n-v}/P_n) \geqq 0$ for $v = 0, 1, \cdots, k$. Hence,

$$\sum_{v=k}^{n}\left(\frac{R_{n-v}Q_v}{P_n}-\frac{R_{n-v-1}Q_v}{P_{n-1}}\right)=\sum_{v=0}^{k-1}Q_v\left(\frac{R_{n-v-1}}{P_{n-1}}-\frac{R_{n-v}}{P_n}\right)\geqq 0$$

and

$$\begin{aligned}\sum_{n=k+N+1}^{M}\left|\sum_{v=k}^{n}\left(\frac{R_{n-v}Q_v}{P_n}-\frac{R_{n-v-1}Q_v}{P_{n-1}}\right)\right| &= \sum_{n=k+N+1}^{M}\sum_{v=k}^{n}\left(\frac{R_{n-v}Q_v}{P_n}-\frac{R_{n-v-1}Q_v}{P_{n-1}}\right)\\ &= \sum_{v=k}^{M}\frac{R_{M-v}Q_v}{P_M}-\sum_{v=k}^{N+k}\frac{R_{N+k-v}Q_v}{P_{N+k}}\\ &\leqq \sum_{v=0}^{M}\frac{|\,R_{M-v}\,|\,Q_v}{P_M}+\sum_{v=0}^{N+k}\frac{|\,R_{N+k-v}\,|\,Q_v}{P_{N+k}}\leqq 2C,\end{aligned} \tag{2.18}$$

independently of M, N and k. Therefore, applying (2.17) and (2.18) to (2.14), we get

$$\sum_{n=0}^{\infty}\left|\sum_{v=k}^{n}\left(\frac{R_{n-v}Q_v}{P_n}-\frac{R_{n-v-1}Q_v}{P_{n-1}}\right)\right|\leqq 2C(N+2)<\infty,$$

independently of k.

(2.19) THEOREM. *If* (1) p_n *is non-negative,* (2) Q_n *is non-negative,* (3) $Q_n/P_n \leqq C < \infty$ *and* (4) *there exists a positive integer* N *such that* R_m *is non-negative and non-increasing whenever* $m > N$, *then* $|\,N_q\,| \subset |\,N_p\,|$.

Proof. We will show that the conditions of Theorem (2.13) are satisfied.

First, $P_n = \sum_{k=0}^{n} p_k \geqq 0$ and $Q_n \geqq 0$. Second,

$$\sum_{v=0}^{n} \left| \frac{R_{n-v} Q_v}{P_n} \right| = \sum_{v=0}^{n-N} \frac{R_{n-v} Q_v}{P_n} + \sum_{v=n-N+1}^{n} \frac{|R_{n-v}| Q_v}{P_n}, \qquad 0 \leqq N \leqq n,$$

$$= 1 - \sum_{v=n-N+1}^{n} \frac{R_{n-v} Q_v}{P_n} + \sum_{v=n-N+1}^{n} \frac{|R_{n-v}| Q_v}{P_n}$$

$$\leqq 1 + 2 \sum_{v=n-N}^{n} \frac{|R_{n-v}| Q_v}{P_n}$$

$$\leqq 1 + 2 \max_{0 \leqq k \leqq N} |R_k| \sum_{v=n-N}^{n} \frac{Q_v}{P_n}$$

$$\leqq 1 + 2C \max_{0 \leqq k \leqq N} |R_k| N \frac{Q_n}{P_n}, \qquad \text{by (1.04)},$$

$$\leqq C' < \infty,$$

since N is finite. Third,

$$\frac{R_{m-1}}{P_{n-1}} - \frac{R_m}{P_n} = R_m \frac{p_n}{P_{n-1} P_n} + \frac{R_{m-1} - R_m}{P_{n-1}}$$

$$\geqq 0 \qquad \text{for } m > N.$$

Thus our theorem is proved.

(2.20) THEOREM. *If* $-1 < \alpha \leqq \beta$, *then* $|C, \alpha| \subset |C, \beta|$.

Proof. Clearly there is no loss of generality in assuming $\gamma = \beta - \alpha \leqq 1$. It is well known that the Cesàro transformation (C, β) is the Nörlund transformation N_p, where $p(x) = (1 - x)^{-\beta}$. Now

$$R(x) = \frac{p(x)}{q(x)} = \frac{(1 - x)^{-\beta}}{(1 - x)^{-\alpha}} = (1 - x)^{-\gamma}.$$

Hence,

$$R_n = \frac{\gamma(\gamma + 1) \cdots (\gamma + n - 1)}{n!}, \tag{2.21}$$

which is non-negative and non-increasing, since $0 \leqq \gamma \leqq 1$.

Case 1. $\beta \geqq 0$. In this case we show that the conditions of Theorem (2.19) are fulfilled.

(1) $p_n = \dfrac{\beta(\beta + 1) \cdots (\beta + n - 1)}{n!} \geqq 0,$ since $\beta \geqq 0$.

(2) $Q_n = \dfrac{(\alpha + 1)(\alpha + 2) \cdots (\alpha + n)}{n!} > 0,$ since $\alpha > -1$.

(3) $\dfrac{Q_n}{P_n} = \dfrac{(\alpha+1)(\alpha+2)\cdots(\alpha+n)}{(\beta+1)(\beta+2)\cdots(\beta+n)} \leqq 1,$ since $\alpha \leqq \beta$.

(4) R_n is non-negative and non-increasing by (2.21).

Case 2. $-1 < \beta < 0$. In this case it will be shown that the conditions of Theorem (2.13) are fulfilled.

(1) $P_n = \dfrac{(\beta+1)(\beta+2)\cdots(\beta+n)}{n!} > 0,$ since $\beta > -1$;

$Q_n = \dfrac{(\alpha+1)(\alpha+2)\cdots(\alpha+n)}{n!} > 0,$ since $\alpha > -1$.

(2) $\displaystyle\sum_{v=0}^{n} \frac{|R_{n-v}|Q_v}{P_n} = \sum_{v=0}^{n} \frac{R_{n-v}Q_v}{P_n},$ since R_n, P_n and $Q_n \geqq 0$,

$= 1,$ by (2.06).

(3) $\dfrac{R_{n-v-1}}{P_{n-1}} - \dfrac{R_{n-v}}{P_n} = \dfrac{R_{n-v-1}}{P_n}\left(\dfrac{P_n}{P_{n-1}} - \dfrac{R_{n-v}}{R_{n-v-1}}\right)$

$= \dfrac{R_{n-v-1}}{P_n}\left(\dfrac{\beta+n}{n} - \dfrac{\gamma+n-v-1}{n-v}\right)$

$= \dfrac{R_{n-v-1}}{n(n-v)P_n}(n\alpha - v\beta + n)$

$\geqq 0,$ since $\alpha + 1 \geqq \beta$.

Thus our theorem is completely proved.

Let us write $\Delta p_n = p_n - p_{n-1}$, and $\Delta^k p_n = \Delta(\Delta^{k-1} p_n)$.

(2.22) THEOREM. *If* (1) $p_n \geqq 0$, (2) $n^k/P_n \leqq C < \infty$, (3) $\Delta^k p_n \geqq 0$ *and* (4) $\Delta^{k+1} p_n \leqq 0$, *then* $|C, k| \subset |N_p|$, $k = 0, 1, 2, \cdots$.

Proof. It will be shown that the conditions of Theorem (2.19) are satisfied.

(1) $p_n \geqq 0,$ by (2.22), (1).

(2) $Q_n = \dfrac{(k+1)(k+2)\cdots(k+n)}{n!} \geqq 0,$ for $k = 0, 1, 2, \cdots$.

(3) $\dfrac{Q_n}{P_n} \leqq C\dfrac{n^k}{P_n} < \infty,$ by (2.22), (2).

(4) $R(x) = \dfrac{p(x)}{q(x)} = (1-x)^k p(x) = 1 + \Delta^k p_1 x + \cdots + \Delta^k p_n x^n + \cdots.$

Therefore, $R_n = \Delta^k p_n \geqq 0$ and $R_n - R_{n-1} = \Delta^{k+1} p_n \leqq 0$.

(2.23) THEOREM. *If* N_q *is the transformation defined by the sequence* $\{q^n\}$ *and* N_p *is the transformation defined by the sequence* $\{p^n\}$, *where* $0 \leqq q \leqq p \leqq 1$, *then* $|N_q| \subset |N_p|$.

Proof. The conditions of Theorem (2.19) are satisfied.

(1) $p_n = p^n \geqq 0.$

(2) $Q_n = \sum_{k=0}^{n} q^k \geqq 0.$

(3) $\frac{Q_n}{P_n} \leqq 1,$ since $q \leqq p.$

(4) $R(x) = \frac{p(x)}{q(x)} = \frac{(1 - px)^{-1}}{(1 - qx)^{-1}}$

$$= 1 + (p - q)x + \cdots + p^{n-1}(p - q)x^n + \cdots.$$

Therefore, $R_n = p^{n-1}(p - q)$, which is clearly non-negative and non-increasing, since $0 \leqq q \leqq p \leqq 1.$

(2.24) THEOREM. *If N_q is the transformation defined by the sequence $\{q^n\}$, $0 \leqq q < 1$, and N_p is the transformation defined by $\{1/(n + 1)\}$, then $|N_q| \subset |N_p|$.*

Proof. The conditions of Theorem (2.19) are satisfied.

(1) $p_n = \frac{1}{n + 1} > 0.$

(2) $Q_n = \sum_{k=0}^{n} q^k > 0.$

(3) $\frac{Q_n}{P_n} = \frac{\sum_{k=0}^{n} q^k}{\sum_{k=0}^{n} (k + 1)^{-1}} \leqq C < \infty.$

(4) $R(x) = \frac{p(x)}{q(x)} = (1 - qx)\left(1 + \frac{x}{2} + \frac{x^2}{3} + \cdots\right).$

Therefore,

$$R_n = \left(\frac{1}{n+1}\right) - \left(\frac{q}{n}\right) \geqq 0 \qquad \text{for } n \geqq \frac{q}{1 - q}$$

and

$$R_n - R_{n+1} = \frac{1}{n+1} - \frac{q}{n} - \frac{1}{n+2} + \frac{q}{n+1} \geqq 0 \qquad \text{for } n \geqq \frac{2q}{1 - q}.$$

(2.25) THEOREM. *If N_q is the transformation defined by $q_n = 1/(n + 1)$, then $|N_q| \subset |C, \alpha|$ for $\alpha > 0$.*

Proof. Due to Theorem (2.20), we can clearly assume $0 < \alpha \leqq 1.$ Again we show that the conditions of Theorem (2.19) are satisfied.

(1) $p_n = \dfrac{\alpha(\alpha+1)\cdots(\alpha+n-1)}{n!} > 0,$ since $\alpha > 0.$

(2) $Q_n = \sum_{k=0}^{n} \dfrac{1}{k+1} > 0.$

(3) $\dfrac{Q_n}{P_n} \leqq C\,\dfrac{\log n}{n^{\alpha}} \leqq C < \infty.$

(4) $R(x) = \dfrac{p(x)}{q(x)} = -\dfrac{x(1-x)^{-\alpha}}{\log(1-x)}.$

Now we observe that

$$\int_0^1 (1-x)^z\,dz = \frac{(1-x)^z}{\log(1-x)}\Bigg]_0^1 = \frac{-x}{\log(1-x)}.$$

Therefore,

$$R(x) = (1-x)^{-\alpha}\int_0^1 (1-x)^z\,dz = \int_0^1 (1-x)^{z-\alpha}\,dz.$$

Hence,

$$\begin{aligned} R_n &= \frac{(-1)^n}{n!}\int_0^1 (z-\alpha)(z-\alpha-1)\cdots(z-\alpha-n+1)\,dz \\ &= \frac{1}{n!}\left\{\int_0^{\alpha}(\alpha-z)(\alpha+1-z)\cdots(\alpha+n-1-z)\,dz\right. \\ &\qquad \left. -\int_{\alpha}^1 (z-\alpha)(\alpha+1-z)\cdots(\alpha+n-1-z)\,dz\right\} \\ &\geqq \frac{1}{n!}\left\{\int_0^{\alpha/2}(\alpha-z)(\alpha+1-z)\cdots(\alpha+n-1-z)\,dz\right. \qquad (2.26) \\ &\qquad \left. -\int_{\alpha}^1 (z-\alpha)(\alpha+1-z)\cdots(\alpha+n-1-z)\,dz\right\} \\ &> \frac{1}{n!}\left\{\frac{\alpha}{2}\cdot\frac{\alpha}{2}\left(1+\frac{\alpha}{2}\right)\cdots\left(n-1+\frac{\alpha}{2}\right)-(1-\alpha)(1-\alpha)\right. \\ &\qquad \left.\times 1\cdot 2\cdots(n-1)\right\} \\ &> 0, \end{aligned}$$

for sufficiently large n. Next we observe that

$$(1-x)\,\frac{p(x)}{q(x)} = (1-x)\int_0^1 (1-x)^{z-\alpha}\,dz = \int_0^1 (1-x)^{z-\alpha+1}\,dz,$$

and, in a manner similar to that above, we find

$$R_n - R_{n-1} < 0, \qquad (2.27)$$

for sufficiently large n. Thus, by (2.26) and (2.27), R_n is non-negative and non-increasing for sufficiently large n, and our theorem is proved.

(2.28) THEOREM. *If* (1) q_n *is non-negative and non-increasing and* (2) q_{n+1}/q_n *is non-decreasing, then* $|N_q| \subset |C, 1|$.

Proof. Again we show that the conditions of Theorem (2.19) are satisfied.

(1) $p_n = 1 > 0$.

(2) $Q_n = \sum_{k=0}^{n} q_k > 0$.

(3) q_n is non-increasing and $q_0 = 1$; therefore, $Q_n \leqq n + 1$. However, $P_n = n + 1$, and so clearly $Q_n/P_n \leqq 1$.

(4) $$R(x) = \frac{p(x)}{q(x)} = \frac{(1-x)^{-1}}{q(x)} = \frac{1}{(1-x)q(x)}$$

$$= \frac{1}{1 - c_1 x - c_2 x^2 - \cdots},$$

where $c_n = q_{n-1} - q_n \geqq 0$. Hence,

$$R_0 = 1 > 0,$$

$$R_1 = c_1 \geqq 0,$$

$$R_2 = R_1 c_1 + c_2 \geqq 0,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots,$$

$$R_n = R_{n-1}c_1 + R_{n-2}c_2 + \cdots + R_1 c_{n-1} + c_n \geqq 0,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots.$$

Also,

$$(1-x)R(x) = \frac{(1-x)(1-x)^{-1}}{q(x)} = \frac{1}{q(x)} = d_0 + d_1 x + d_2 x^2 + \cdots.$$

Kaluza [9] proved that if $f(x) = a_0 + a_1 x + \cdots + a_n x^n + \cdots$ and $1/f(x) = b_0 + b_1 x + \cdots + b_n x^n + \cdots$, where a_n is non-negative and a_{n+1}/a_n is non-decreasing, then b_n is non-positive. Thus it follows that $R_n - R_{n-1} = d_n \leqq 0$; hence R_n is non-negative and non-increasing.

(2.29) THEOREM. *If* (1) p_n *is non-negative and non-decreasing and* (2) p_{n+1}/p_n *is non-increasing, then* $|C, 1| \subset |N_p|$.

Proof. We will show that the conditions of Theorem (2.12) are satisfied. First, observe that

$$R(x) = \frac{p(x)}{(1-x)^{-1}} = (1-x)p(x); \tag{2.30}$$

therefore,

$$R_n = p_n - p_{n-1}. \tag{2.31}$$

Now (1)

$$\sum_{v=k}^{n}\left(\frac{R_{n-v}Q_v}{P_n}-\frac{R_{n-v-1}Q_v}{P_{n-1}}\right)=\sum_{v=k}^{n}\left\{\frac{(p_{n-v}-p_{n-v-1})(v+1)}{P_n}\right.$$

$$\left.-\frac{(p_{n-v-1}-p_{n-v-2})(v+1)}{P_{n-1}}\right\},\quad \text{by (2.31)},$$

(2.32)

$$=\frac{kp_{n-k}+P_{n-k}}{P_n}-\frac{kp_{n-k-1}+P_{n-k-1}}{P_{n-1}}$$

$$=k\left(\frac{p_{n-k}}{P_n}-\frac{p_{n-k-1}}{P_{n-1}}\right)+\frac{P_{n-k}}{P_n}-\frac{P_{n-k-1}}{P_{n-1}}.$$

From (2.29), (2), it is easily deducible that

$$\frac{P_{n-k}}{P_n}-\frac{P_{n-k-1}}{P_{n-1}}\geqq 0. \tag{2.33}$$

Thus, if

$$\frac{p_{n-k}}{P_n}-\frac{p_{n-k-1}}{P_{n-1}}\geqq 0,$$

then, due to (2.33), the expression on the left hand side of (2.32) is non-negative. Suppose on the other hand that

$$\frac{p_{n-k}}{P_n}-\frac{p_{n-k-1}}{P_{n-1}}<0.$$

Then

$$\frac{p_{n-k}}{P_n}-\frac{p_{n-k-1}}{P_{n-1}}\geqq\frac{p_{n-k-1}}{P_n}\left(\frac{p_{n-m}}{p_{n-m-1}}-\frac{P_n}{P_{n-1}}\right),\quad \text{for } 0\leqq m\leqq k, \text{ due to (2.29), (2),}$$

$$\geqq\frac{p_{n-m}}{P_n}-\frac{p_{n-m-1}}{P_{n-1}}.$$

Therefore,

$$k\left(\frac{p_{n-k}}{P_n}-\frac{p_{n-k-1}}{P_{n-1}}\right)\geqq\sum_{m=k-1}^{0}\left(\frac{p_{n-m}}{P_n}-\frac{p_{n-m-1}}{P_{n-1}}\right),$$

and, substituting in (2.32), we get

$$\sum_{v=k}^{n}\left(\frac{R_{n-v}Q_v}{P_n}-\frac{R_{n-v-1}Q_v}{P_{n-1}}\right)\geqq\sum_{m=k-1}^{0}\frac{p_{n-m}}{P_n}+\frac{P_{n-k}}{P_n}-\sum_{m=k-1}^{0}\frac{p_{n-m-1}}{P_{n-1}}-\frac{P_{n-k-1}}{P_{n-1}}$$

$$=0.$$

(2) $R_n=p_n-p_{n-1}$, by (2.31),

$\geqq 0$, by (2.29), (1).

Therefore,

$$\sum_{v=0}^{n}\left|\frac{R_{n-v}Q_v}{P_n}\right|=\sum_{v=0}^{n}\frac{R_{n-v}Q_v}{P_n}=1,\qquad \text{by (2.06)}.$$

Thus our theorem is established.

3. **Abel's transformation.** The formula

$$\sum_{k=m}^{n} u_k v_k = \sum_{k=m}^{n-1} U_k(v_k - v_{k+1}) - U_{m-1}v_m + U_n v_n, \tag{3.01}$$

where $0 \leqq m \leqq n$, $U_k = u_0 + u_1 + \cdots + u_k$, if $k \geqq 0$, $U_{-1} = 0$, which can be verified, is known as Abel's transformation, and will be used extensively in what follows.

(3.02) COROLLARY. *If $v_m, v_{m+1}, \cdots, v_n$ are non-negative and non-increasing, the left hand side of* (3.01) *does not exceed* $2v_m \max_{m-1 \leq k \leq n} |U_k|$ *in absolute value. In fact,*

$$\left|\sum_{k=m}^{n} u_k v_k\right| \leqq \max |U_k| \left\{\sum_{k=m}^{n-1}(v_k - v_{k+1}) + v_m + v_n\right\}$$
$$= 2v_m \max |U_k|.$$

4. **Absolute Abel summability.** The series $\sum_{n=0}^{\infty} u_n$ is said to be summable by Abel's method, or summable A, to sum U if the expression

$$f(x) = u_0 + u_1 x + \cdots + u_n x^n + \cdots \tag{4.01}$$

is convergent for $|x| < 1$, and

$$\lim_{x \to 1} f(x) = \lim_{x \to 1} \sum_{k=0}^{\infty} u_k x^k = \lim_{x \to 1} (1 - x) \sum_{k=0}^{\infty} U_k x^k = U,$$

where x tends to 1 along the real axis.

Following the definition of Whittaker [10], we will say that the series $\sum_{n=0}^{\infty} u_n$ is absolutely Abel summable, or summable $|A|$, if $f(x)$ is of bounded variation on $[0, 1]$, that is, if $\int_0^1 |f'(x)|\, dx$ exists.

Fekete [3] showed that $\sum_{n=0}^{\infty} u_n$ is summable $|A|$ if it is summable $|C, r|$ for r a positive integer.

Bosanquet [2] extended Fekete's result to include non-integral values of r.

We now prove the following

(4.02) . THEOREM. *Let N_p be a regular Nörlund transformation and let us write*

$$\phi_n(x) = \sum_{k=0}^{n} P_k x^k \Big/ \sum_{k=0}^{\infty} P_k x^k.$$

If (1) $\sum_{n=0}^{\infty} u_n$ *is summable* $|N_p|$ *and* (2) *the sequence* $\{\phi_n(x)\}$ *is uniformly of bounded variation, then* $\sum_{n=0}^{\infty} u_n$ *is summable* $|A|$.

Proof. We will write $P(x) = \sum_{n=0}^{\infty} P_n x^n$ and $R(x) = \sum_{n=0}^{\infty} P_n t_n x^n$. Clearly, $R(x) = \sum_{n=0}^{\infty} \sum_{v=0}^{n} P_{n-v} u_v x^n = f(x)P(x)$; hence, $f(x) = R(x)/P(x)$. Since N_p is regular, we have $\lim_{n \to \infty} (P_{n-1}/P_n) = 1$, so $P(x)$ has a radius of convergence 1. By hypothesis, $\sum_{n=0}^{\infty} | t_n - t_{n-1} | \leqq C < \infty$; hence, t_n approaches a finite limit, and thus $R(x)$ has a radius of convergence 1. Therefore, we can write

$$
\begin{aligned}
f'(x) &= \frac{P(x)R'(x) - P'(x)R(x)}{P^2(x)} \\
&= \frac{P(x) \sum_{n=0}^{\infty} nP_n t_n x^{n-1} - P'(x) \sum_{n=0}^{\infty} P_n t_n x^n}{P^2(x)} \\
&= \frac{P(x) \sum_{n=0}^{\infty} \sum_{k=1}^{n} kP_k x^{k-1}(t_n - t_{n+1}) - P'(x) \sum_{n=0}^{\infty} \sum_{k=0}^{n} P_k x^k (t_n - t_{n+1})}{P^2(x)}, \\
&\qquad\qquad \text{by (3.01) (Abel's transformation),} \\
&= \sum_{n=0}^{\infty} \left\{ \frac{P(x)\, d(P_0 + \cdots + P_n x^n)/dx - P'(x)(P_0 + \cdots + P_n x^n)}{P^2(x)} \right\} (t_n - t_{n+1}) \\
&= \sum_{n=0}^{\infty} \phi_n'(x)(t_n - t_{n+1}).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\int_0^1 | f'(x) |\, dx &\leqq \sum_{n=0}^{\infty} \int_0^1 | \phi_n'(x) |\, dx\, | t_n - t_{n+1} | \\
&\leqq C \sum_{n=0}^{\infty} | t_n - t_{n+1} |, && \text{by (4.02), (2),} \\
&< \infty, && \text{by (4.02), (1),}
\end{aligned}
$$

which is the condition that $\sum_{n=0}^{\infty} u_n$ be summable $| A |$.

(4.03) THEOREM. *If* (1) N_p *is regular,* (2) $P_n \geqq 0$ *and* (3) $\sum_{n=0}^{\infty} u_n$ *is summable* $| N_p |$, *then* $\sum_{n=0}^{\infty} u_n$ *is summable* $| A |$.

Proof. It is clear from (4.02) that we only need to show that $\{\phi_n(x)\}$ is uniformly of bounded variation. Now

$$
\begin{aligned}
\phi_n'(x) &= \frac{d}{dx} \left\{ \sum_{k=0}^{n} P_k x^k \Big/ \sum_{k=0}^{\infty} P_k x^k \right\} \\
&= \sum_{k=0}^{\infty} V_k x^k \Big/ \left(\sum_{k=0}^{\infty} P_k x^k \right)^2,
\end{aligned}
$$

where

$$V_k = P_1P_k + 2P_2P_{k-1} + \cdots + nP_nP_{k-n+1} - (k+1)P_0P_{k+1} + kP_1P_k + \cdots + (k-n+1)P_nP_{k-n+1},$$

it being understood that $P_r = 0$ when r is negative. If we consider separately the cases $0 \leqq k \leqq n$, $n < k \leqq 2n$ and $k > 2n$, it is easily seen that in all cases $V_k \leqq 0$. Therefore $\phi_n'(x) \leqq 0$ for all n and $0 \leqq x < 1$. Hence

$$\int_0^1 |\phi_n'(x)|\,dx = -\int_0^1 \phi_n'(x)\,dx = -\phi_n(x)\Big]_0^1 = 1 \qquad \text{for all } n.$$

Thus our theorem is established.

5. **Fourier series.** In this section we will be concerned with a function $f(x)$ of period 2π and belonging to some class L^q, $q \geqq 1$. If the Fourier series $S(f)$ and the conjugate series $\tilde{S}(f)$ of the function $f(x)$ are respectively

$$\tfrac{1}{2}a_0 + \sum_{k=1}^{\infty}(a_k \cos kx + b_k \sin kx), \qquad \sum_{k=1}^{\infty}(a_k \sin kx - b_k \cos kx), \tag{5.01}$$

we can write for the n-th partial sums $S_n(f, x)$ and $\tilde{S}_n(f, x)$ of these series

$$S_n(x) - f(x) = \frac{1}{\pi}\int_0^{\pi} \phi(t)D_n(t)\,dt, \qquad \tilde{S}_n(x) = -\frac{1}{\pi}\int_0^{\pi} \psi(t)\tilde{D}_n(t)\,dt, \tag{5.02}$$

where $\phi(t) = \phi_x(t) = f(x+t) + f(x-t) - 2f(x)$, $\psi(t) = \psi_x(t) = f(x+t) - f(x-t)$,

$$D_n(t) = \frac{\sin\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}, \qquad \tilde{D}_n(t) = \frac{\cos\frac{t}{2} - \cos\left(n+\frac{1}{2}\right)t}{2\sin\frac{t}{2}}.$$

Bernstein [1] showed that, if $f(x)$ belongs to Lip α, the Fourier series of $f(x)$ converges absolutely for all values of x when $\alpha > \frac{1}{2}$, but not necessarily when $\alpha \leqq \frac{1}{2}$. Later Hyslop [5] extended Bernstein's theorem by showing that, if $f(x)$ belongs to Lip α, where $0 < \alpha \leqq \frac{1}{2}$, the Fourier series of $f(x)$ is summable $|C, \beta|$ for all values of x if $\alpha + \beta > \frac{1}{2}$.

Our object will be to extend the result of Hyslop to the case of summability $|N_p|$ for certain types of Nörlund means. Analogous theorems for the conjugate series will be established at the same time, and certain other results obtained.

(5.03) **Notation and lemmas.** We will write $p(y) = p_{[y]}$, and $P(y) = P_{[y]}$, where $[y]$ as usual denotes the greatest integer less than y.

$$(5.04)\qquad t_n = \sum_{\nu=0}^{n} \frac{p_\nu S_{n-\nu}}{P_n}, \qquad \tilde{t}_n = \sum_{\nu=0}^{n} \frac{p_\nu \tilde{S}_{n-\nu}}{P_n}.$$

$$(5.05)\qquad \Phi(h) = \int_0^h |\phi(t)|\, P(t^{-1})\, dt, \qquad \Psi(h) = \int_0^h |\psi(t)|\, P(t^{-1})\, dt.$$

$$(5.06)\qquad \alpha(t) = \sum_{k=0}^{\infty} p_k \cos kt.$$

$$(5.07)\qquad \beta(t) = \sum_{k=0}^{\infty} p_k \sin kt.$$

$$(5.08)\qquad \alpha_n = \int_0^\pi \phi(t)\alpha(t) \cos nt\, dt, \qquad \tilde{\alpha}_n = \int_0^\pi \psi(t)\alpha(t) \sin nt\, dt.$$

$$(5.09)\qquad \beta_n = \int_0^\pi \phi(t)\beta(t) \sin nt\, dt, \qquad \tilde{\beta}_n = \int_0^\pi \psi(t)\beta(t) \cos nt\, dt.$$

Furthermore, A will denote an absolute positive constant, and we will write $A + A = A$ and $A \cdot A = A$. By $f_n \simeq g_n$, it will be meant that there exist two constants $A_1 > 0$ and $A_2 > 0$, such that $A_1 \leqq f_n/g_n \leqq A_2$ for n sufficiently large.

(5.10) LEMMA. *If p_n is non-negative and non-increasing, then $t^{-1}p(t^{-1}) \leqq P(t^{-1})$.*

Proof. $P_n = p_0 + p_1 + \cdots + p_n \geqq (n+1)p_n$. Therefore, $t^{-1}p(t^{-1}) \leqq ([t^{-1}] + 1)p_{[t^{-1}]} \leqq P_{[t^{-1}]} \leqq P(t^{-1})$.

(5.11) LEMMA.[2] *If p_n is non-negative and non-increasing, then, for $0 \leqq a \leqq b \leqq \infty$, $0 \leqq t \leqq \pi$ and any n, we have*

$$\left|\sum_{k=a}^{b} p_k e^{i(n-k)t}\right| \leqq \begin{cases} P(t^{-1}), & \text{for any } a, \\ At^{-1}p_a, & \text{for } a \geqq [t^{-1}]. \end{cases}$$

Proof. Let $\tau = [t^{-1}]$. Then

$$\left|\sum_{k=a}^{b} p_k e^{i(n-k)t}\right| = \left|e^{int}\sum_{k=a}^{b} p_k e^{-ikt}\right|$$

$$\leqq \left|\sum_{k=a}^{\tau-1} p_k e^{-ikt}\right| + \left|\sum_{k=\tau}^{b} p_k e^{-ikt}\right|;$$

but

$$\left|\sum_{k=a}^{\tau-1} p_k e^{-ikt}\right| \leqq \sum_{k=a}^{\tau-1} p_k \leqq P_\tau \leqq P(t^{-1}),$$

and, by (3.02),

$$\left|\sum_{k=\tau}^{b} p_k e^{-ikt}\right| \leqq 2\, p_\tau \max_{\tau+1 \leq k \leq b} \left|\frac{1 - e^{-i(k+1)t}}{1 - e^{-it}}\right|$$

$$\leqq 4p_\tau \left|\frac{e^{it/2}}{e^{it/2} - e^{-it/2}}\right|$$

$$\leqq 2p_\tau(1/\sin \tfrac{1}{2}t)$$

$$\leqq At^{-1}p(t^{-1}).$$

[2] This lemma is due to Tamarkin and Hille.

The lemma follows immediately, since, by Lemma (5.10), $t^{-1}p(t^{-1}) \leqq P(t^{-1})$, and, in case $a \geqq [t^{-1}]$, we would have

$$\left|\sum_{k=a}^{b} p_k e^{-ikt}\right| \leqq 2p_a \max_{a\leq k\leq b}\left|\frac{1-e^{-i(k+1)t}}{1-e^{-it}}\right| \leqq At^{-1}p_a.$$

(5.12) LEMMA. *If p_n is non-negative and non-increasing, then*

$$\frac{(P_0+P_1+\cdots+P_n)}{P_n}\,p_n \leqq P(t^{-1}) \qquad \text{for } t \leqq \frac{1}{n}.$$

Proof.

$$\frac{(P_0+\cdots+P_n)}{P_n}\,p_n \leqq (n+1)p_n \leqq P_n \leqq P(t^{-1}) \qquad \text{for } t \leqq \frac{1}{n}.$$

(5.13) LEMMA. *If (1) p_n is non-negative and non-increasing and (2) $p_n - p_{n+1}$ is non-increasing, then*

$$\frac{(n+1)^2(p_n-p_{n+1})}{P_n} \leqq A.$$

Proof.

$$\begin{aligned}(n+1)^2(p_n-p_{n+1}) &\leqq A\left(\frac{n(n+1)}{2}\right)(p_n-p_{n+1})\\ &= A\sum_{k=0}^{n}(k+1)(p_n-p_{n+1})\\ &\leqq A\sum_{k=0}^{n}(k+1)(p_k-p_{k+1}), \quad \text{by (5.13), (2),}\\ &= A\{P_n-(n+1)\,p_{n+1}\}\\ &\leqq AP_n.\end{aligned}$$

(5.14) LEMMA. *If p_n is non-negative and non-increasing, then*

$$\sum_{k=n}^{\infty}\frac{k(p_k-p_{k+1})}{P_kP_{k-1}} \leqq \frac{A}{P_{n-1}}.$$

Proof.

$$\begin{aligned}\sum_{k=n}^{\infty}\frac{k(p_k-p_{k+1})}{P_kP_{k-1}} &= \frac{n(p_n-p_{n+1})}{P_nP_{n-1}} + \frac{(n+1)(p_{n+1}-p_{n+2})}{P_{n+1}P_n} + \cdots\\ &= \frac{np_n}{P_nP_{n-1}} + \sum_{k=n}^{\infty}\frac{p_{k+1}}{P_k}\left(\frac{k+1}{P_{k+1}} - \frac{k}{P_{k-1}}\right)\\ &= \frac{np_n}{P_nP_{n-1}} + \sum_{k=n}^{\infty}\frac{kp_{k+1}}{P_k}\left(\frac{1}{P_{k+1}} - \frac{1}{P_{k-1}}\right) + \sum_{k=n}^{\infty}\frac{p_{k+1}}{P_kP_{k+1}}\\ &\leqq \frac{np_n}{P_nP_{n-1}} + \sum_{k=n}^{\infty}\frac{p_{k+1}}{P_kP_{k+1}} \leqq \frac{A}{P_{n-1}} + \frac{1}{P_n} \leqq \frac{A}{P_{n-1}}.\end{aligned}$$

(5.15) LEMMA. *If p_n is non-negative and non-increasing, then $n^{-1}P_n \leqq tP(t^{-1})$ for $1/n \leqq t \leqq \pi$.*

Proof.

$$\frac{P_n}{n} - \frac{P_{n+1}}{n+1} = \frac{P_n - np_{n+1}}{n(n+1)} \geqq 0.$$

Hence P_n/n decreases as n increases and the result follows.

(5.16) LEMMA. *Let N_p be a given Nörlund transformation with p_n non-negative and non-increasing, $\lim_{n\to\infty} p_n = 0$ and $|\Delta p_n|$ non-increasing. Suppose also that $\phi(t)P(t^{-1})$ belongs to class L^q, $q \geqq 1$. Then*

$$(5.17) \qquad \sum_{n=1}^{\infty} |t_n - t_{n-1}| \leqq A\left\{1 + \sum_{n=1}^{\infty} \frac{\Phi(n^{-1})}{P_{n-1}} + \sum_{n=1}^{\infty} \frac{|\alpha_n| + |\beta_n|}{P_{n-1}}\right\}.$$

Proof.

$$\begin{aligned}
\pi(t_n - t_{n-1}) &= \int_0^{\pi} \phi(t) \sum_{k=0}^{n} \left(\frac{p_{n-k}}{P_n} - \frac{p_{n-k-1}}{P_{n-1}}\right) D_k(t)\, dt \\
&= \int_0^{\pi} \phi(t) \sum_{k=1}^{n} \left(\frac{P_{n-k}}{P_n} - \frac{P_{n-k-1}}{P_{n-1}}\right) \cos kt\, dt \\
&= \int_0^{\pi} \phi(t) \sum_{k=0}^{n-1} \left(\frac{P_k}{P_n} - \frac{P_{k-1}}{P_{n-1}}\right) \cos (n-k)t\, dt \\
&= \frac{1}{P_n P_{n-1}} \int_0^{\pi} \phi(t) \sum_{k=0}^{n-1} (P_k P_{n-1} - P_n P_{k-1}) \cos (n-k)t\, dt.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\pi(t_n - t_{n-1}) &= \frac{1}{P_n P_{n-1}} \int_0^{\pi} \phi(t) \sum_{k=0}^{n-1} (p_k P_n - p_n P_k) \cos (n-k)t\, dt \\
&= \frac{1}{P_n P_{n-1}} \int_0^{\pi} \phi(t) \left\{\sum_{k=0}^{\infty} p_k P_n - \sum_{k=n}^{\infty} p_k P_n - \sum_{k=0}^{n-1} p_n P_k\right\} \cos (n-k)t\, dt.
\end{aligned}$$

Hence,

$$\begin{aligned}
\pi|t_n - t_{n-1}| &\leqq \frac{1}{P_{n-1}} \left| \int_0^{\pi} \phi(t) \sum_{k=0}^{\infty} p_k \cos (n-k)t\, dt \right| \\
&\quad + \frac{1}{P_{n-1}} \left| \int_0^{1/n} \phi(t) \sum_{k=n}^{\infty} p_k \cos (n-k)t\, dt \right| \\
&\quad + \frac{p_n}{P_n P_{n-1}} \left| \int_0^{1/n} \phi(t) \sum_{k=0}^{n-1} P_k \cos (n-k)t\, dt \right| \\
&\quad + \frac{1}{P_{n-1}} \left| \int_{1/n}^{\pi} \phi(t) \left\{\sum_{k=n}^{\infty} p_k \cos (n-k)t + \sum_{k=0}^{n-1} \frac{p_n}{P_n} P_k \cos (n-k)t\right\} dt \right| \\
&= I_1(n) + I_2(n) + I_3(n) + I_4(n).
\end{aligned}$$

Now

$$I_1(n) = \frac{1}{P_{n-1}}\left|\int_0^\pi \phi(t)\sum_{k=0}^{\infty} p_k \cos(n-k)t\,dt\right|$$

$$= \frac{1}{P_{n-1}}\left|\int_0^\pi \phi(t)\left\{\sum_{k=0}^{\infty} p_k \cos kt \cos nt + \sum_{k=0}^{\infty} p_k \sin kt \sin nt\right\}dt\right|$$

$$= \frac{1}{P_{n-1}}\left|\int_0^\pi \phi(t)\alpha(t)\cos nt\,dt\right| + \frac{1}{P_{n-1}}\left|\int_0^\pi \phi(t)\beta(t)\sin nt\,dt\right|$$

$$= \frac{|\alpha_n| + |\beta_n|}{P_{n-1}}.$$

$$I_2(n) = \frac{1}{P_{n-1}}\left|\int_0^{1/n} \phi(t)\sum_{k=n}^{\infty} p_k \cos(n-k)t\,dt\right|$$

$$\leqq \frac{A}{P_{n-1}}\int_0^{1/n} |\phi(t)|\,P(t^{-1})\,dt, \qquad \text{by Lemma (5.11),}$$

$$= A\,\frac{\Phi(n^{-1})}{P_{n-1}}, \qquad \text{by (5.05).}$$

$$I_3(n) = \frac{p_n}{P_n P_{n-1}}\left|\int_0^{1/n} \phi(t)\sum_{k=0}^{n-1} P_k \cos(n-k)t\,dt\right|$$

$$\leqq \frac{(P_0 + \cdots + P_n)p_n}{P_n P_{n-1}}\int_0^{1/n} |\phi(t)|\,dt$$

$$\leqq \frac{1}{P_{n-1}}\int_0^{1/n} |\phi(t)|\,P(t^{-1})\,dt, \qquad \text{by Lemma (5.12),}$$

$$= \Phi(n^{-1})/P_{n-1}.$$

$$I_4(n) = \frac{1}{P_{n-1}}\left|\int_{1/n}^\pi \phi(t)\left\{\sum_{k=n}^{\infty} p_k \cos(n-k)t + \sum_{k=0}^{n-1}\frac{p_n}{P_n}P_k \cos(n-k)t\right\}dt\right|.$$

Applying Abel's transformation we obtain

$$\sum_{k=n}^{\infty} p_k \cos(n-k)t = p_n \cos 0 + p_{n+1}\cos t + p_{n+2}\cos 2t + \cdots$$

$$= \frac{p_n}{2} + (p_n - p_{n+1})(\tfrac{1}{2}) + (p_{n+1} - p_{n+2})(\tfrac{1}{2} + \cos t) + \cdots$$

$$= \frac{p_n}{2} + \sum_{k=n}^{\infty}(p_k - p_{k+1})\frac{\sin(n-k+\frac{1}{2})t}{2\sin\frac{1}{2}t}.$$

$$\sum_{k=0}^{n-1} P_k \cos(n-k)t = P_0 \cos nt + P_1 \cos(n-1)t + \cdots + P_{n-1}\cos t$$

$$= (P_{n-1} - P_{n-2})(\tfrac{1}{2} + \cos t) + \cdots + (P_1 - P_0)$$

$$\times(\tfrac{1}{2} + \cos t + \cdots + \cos(n-1)t)$$

$$+ P_0(\tfrac{1}{2} + \cos t + \cdots + \cos nt) - \tfrac{1}{2}P_{n-1}$$

$$= \sum_{k=0}^{n-1} p_k \frac{\sin(n-k+\frac{1}{2})t}{2\sin\frac{1}{2}t} - \tfrac{1}{2}P_{n-1}.$$

Hence,

$$I_4(n) \leqq \frac{1}{P_{n-1}} \left| \int_{1/n}^{\pi} \frac{\phi(t)}{2 \sin \frac{1}{2}t} \sum_{k=n}^{\infty} (p_k - p_{k+1}) \sin (n - k + \tfrac{1}{2})t \, dt \right|$$
$$+ \frac{p_n}{P_n P_{n-1}} \left| \int_{1/n}^{\pi} \frac{\phi(t)}{2 \sin \frac{1}{2}t} \sum_{k=0}^{n-1} p_k \sin (n - k + \tfrac{1}{2})t \, dt \right|$$
$$+ \frac{1}{2} \frac{p_n}{P_{n-1}} \left(1 - \frac{P_{n-1}}{P_n}\right) \left| \int_{1/n}^{\pi} \phi(t) \, dt \right|$$
$$= I_{4,1}(n) + I_{4,2}(n) + I_{4,3}(n).$$

We observe that

$$\sum_{k=n}^{\infty} (p_k - p_{k+1}) \sin (n - k + \tfrac{1}{2})t = \Im \left\{ \sum_{k=n}^{\infty} (p_k - p_{k+1}) \exp (i(n - k + \tfrac{1}{2})t) \right\},$$

and $n \geqq t^{-1}$ in the interval $[1/n, \pi]$, so that, by Lemma (5.11), we have

$$I_{4,1}(n) \leqq A \frac{(p_n - p_{n+1})}{P_{n-1}} \int_{1/n}^{\pi} \frac{|\phi(t)|}{\sin t} t^{-1} \, dt$$
$$= A \frac{n(p_n - p_{n+1})}{P_n P_{n-1}} \int_{1/n}^{\pi} \frac{|\phi(t)|}{\sin t} t^{-1} \frac{P_n}{n} \, dt$$
$$\leqq A \frac{n(p_n - p_{n+1})}{P_n P_{n-1}} \int_{1/n}^{\pi} \frac{|\phi(t)|}{\sin t} P(t^{-1}) \, dt, \qquad \text{by Lemma (5.15),}$$
$$\leqq A \frac{n(p_n - p_{n+1})}{P_n P_{n-1}} \int_{1/n}^{\pi} |\phi(t)| P(t^{-1}) t^{-1} \, dt.$$

Applying Lemma (5.11) to $I_{4,2}(n)$, we get

$$I_{4,2}(n) \leqq (Ap_n / P_n P_{n-1}) \int_{1/n}^{\pi} |\phi(t)| P(t^{-1}) t^{-1} \, dt.$$

On integrating by parts, we see that

$$\int_{1/n}^{\pi} |\phi(t)| P(t^{-1}) t^{-1} \, dt = t^{-1} \Phi(t) \Big]_{1/n}^{\pi} + \int_{1/n}^{\pi} \Phi(t) t^{-2} \, dt$$
$$\leqq \pi^{-1} \Phi(\pi) + n\Phi(n^{-1}) + \sum_{k=1}^{n} \Phi(k^{-1}) \int_{1/(k+1)}^{1/k} t^2 \, dt + A$$
$$\leqq A + n\Phi(n^{-1}) + \sum_{k=1}^{n} \Phi(k^{-1}).$$

Therefore,

$$\sum_{n=1}^{\infty} I_{4,1}(n) \leqq A \left\{ \sum_{n=1}^{\infty} \frac{n(p_n - p_{n+1})}{P_n P_{n-1}} + \sum_{n=1}^{\infty} \frac{n^2(p_n - p_{n+1})}{P_n P_{n-1}} \Phi(n^{-1}) \right.$$
$$\left. + \sum_{n=1}^{\infty} \frac{n(p_n - p_{n+1})}{P_n P_{n-1}} \sum_{k=1}^{n} \Phi(k^{-1}) \right\}.$$

However,

$$\sum_{n=1}^{\infty} \frac{n(p_n - p_{n+1})}{P_n P_{n-1}} \leqq A \qquad \text{by Lemma (5.14) with } n = 1;$$

$$\frac{n^2(p_n - p_{n+1})}{P_n P_{n-1}} \leqq \frac{(n+1)^2(p_n - p_{n+1})}{P_n P_{n-1}} \leqq \frac{A}{P_{n-1}} \qquad \text{by Lemma (5.13);}$$

$$\sum_{n=1}^{\infty} \frac{n(p_n - p_{n+1})}{P_n P_{n-1}} \sum_{k=1}^{n} \Phi(k^{-1}) = \sum_{n=1}^{\infty} \Phi(n^{-1}) \sum_{k=n}^{\infty} \frac{k(p_k - p_{k+1})}{P_k P_{k-1}}$$

$$\leqq A \sum_{n=1}^{\infty} \frac{\Phi(n^{-1})}{P_{n-1}} \qquad \text{by Lemma (5.14).}$$

Therefore,

$$\sum_{n=1}^{\infty} I_{4,1}(n) \leqq A \left\{1 + \sum_{n=1}^{\infty} \frac{\Phi(n^{-1})}{P_{n-1}}\right\}.$$

Also,

$$\sum_{n=1}^{\infty} I_{4,2}(n) \leqq A \left\{\sum_{n=1}^{\infty} \frac{p_n}{P_n P_{n-1}} + \sum_{n=1}^{\infty} \frac{np_n}{P_n P_{n-1}} \Phi(n^{-1}) + \sum_{n=1}^{\infty} \frac{p_n}{P_n P_{n-1}} \sum_{k=1}^{n} \Phi(k^{-1})\right\}.$$

However,

$$\sum_{n=1}^{\infty} \frac{p_n}{P_n P_{n-1}} = \sum_{n=1}^{\infty} \left(\frac{1}{P_{n-1}} - \frac{1}{P_n}\right) = \frac{1}{P_0} - \frac{1}{P_\infty} \leqq 1;$$

$$\frac{np_n}{P_n P_{n-1}} \leqq \frac{1}{P_{n-1}},$$

since $np_n \leqq (n+1)p_n \leqq P_n$; and

$$\sum_{n=1}^{\infty} \frac{p_n}{P_n P_{n-1}} \sum_{k=1}^{n} \Phi(k^{-1}) = \sum_{n=1}^{\infty} \Phi(n^{-1}) \sum_{k=n}^{\infty} \frac{p_k}{P_k P_{k-1}} = \sum_{n=1}^{\infty} \frac{\Phi(n^{-1})}{P_{n-1}}.$$

Therefore,

$$\sum_{n=1}^{\infty} I_{4,2}(n) \leqq A \left\{1 + \sum_{n=1}^{\infty} \frac{\Phi(n^{-1})}{P_{n-1}}\right\}.$$

Finally,

$$I_{4,3}(n) = \frac{1}{2} \frac{p_n^2}{P_n P_{n-1}} \int_0^{\pi} |\phi(t)|\, dt.$$

Now $\phi(t)$ is integrable, $(n+1)p_n \leqq P_n$, and, due to regularity, $(n+1)p_n \leqq AP_{n-1}$. Therefore,

$$I_{4,3}(n) \leqq \frac{A}{(n+1)^2}, \qquad \sum_{n=1}^{\infty} I_{4,3}(n) \leqq A.$$

Hence (5.17) is established and the lemma is proved.

(5.18) LEMMA. *Consider the Nörlund transformation N_p with p_n non-negative and non-increasing,* $\lim_{n\to\infty} p_n = 0$ *and* $|\Delta p_n|$ *non-increasing. Suppose also that* $\psi(t)P(t^{-1})$ *belongs to class* L^q, $q \geqq 1$. *Then*

$$\sum_{n=1}^{\infty} |\bar{\imath}_n - \bar{\imath}_{n-1}| \leqq A\left\{1 + \sum_{n=1}^{\infty} \frac{\Psi(n^{-1})}{P_{n-1}} + \sum_{n=1}^{\infty} \frac{|\tilde{\alpha}_n| + |\tilde{\beta}_n|}{P_{n-1}}\right\}. \tag{5.19}$$

Proof.

$$\pi(\bar{\imath}_n - \bar{\imath}_{n-1}) = -\int_0^{\pi} \psi(t) \sum_{k=0}^{n} \left(\frac{p_{n-k}}{P_n} - \frac{p_{n-k-1}}{P_{n-1}}\right) \tilde{D}_k(t)\, dt.$$

Going through the same steps as in Lemma (5.16) with $\psi(t)$ in place of $\phi(t)$ and $\tilde{D}_k(t)$ in place of $D_k(t)$, it follows immediately that

$$\begin{aligned}
\pi|\bar{\imath}_n - \bar{\imath}_{n-1}| &\leqq \frac{1}{P_{n-1}}\left|\int_0^{\pi} \psi(t) \sum_{k=0}^{\infty} p_k \sin(n-k)t\, dt\right| \\
&\quad + \frac{1}{P_{n-1}}\left|\int_0^{\pi} \psi(t) \sum_{k=n}^{\infty} p_k \sin(n-k)t\, dt\right| \\
&\quad + \frac{p_n}{P_n P_{n-1}}\left|\int_0^{1/n} \psi(t) \sum_{k=0}^{n-1} P_k \sin(n-k)t\, dt\right| \\
&\quad + \frac{1}{P_{n-1}}\left|\int_{1/n}^{\pi} \psi(t)\left\{\sum_{k=n}^{\infty} p_k \sin(n-k)t + \sum_{k=0}^{n-1} \frac{p_n}{P_n} P_k \sin(n-k)t\right\} dt\right| \\
&= I_1(n) + I_2(n) + I_3(n) + I_4(n).
\end{aligned}$$

$$\begin{aligned}
I_1(n) &= \frac{1}{P_{n-1}}\left|\int_0^{\pi} \psi(t)\left\{\sum_{k=0}^{\infty} p_k \cos kt \sin nt - \sum_{k=0}^{\infty} p_k \sin kt \cos nt\right\} dt\right| \\
&\leqq \frac{|\tilde{\alpha}_n| + |\tilde{\beta}_n|}{P_{n-1}}.
\end{aligned}$$

$$\begin{aligned}
I_2(n) &\leqq \frac{A}{P_{n-1}} \int_0^{1/n} |\psi(t)|\, P(t^{-1})\, dt, \qquad \text{by Lemmas (5.10) and (5.11),} \\
&= A\frac{\Psi(n^{-1})}{P_{n-1}}.
\end{aligned}$$

$$\begin{aligned}
I_3(n) &\leqq \frac{(P_0 + \cdots + P_n)}{P_n P_{n-1}} p_n \int_0^{1/n} |\psi(t)|\, dt \leqq \frac{1}{P_{n-1}} \int_0^{1/n} \psi(t) P(t^{-1})\, dt, \\
&\qquad \text{by Lemma (5.12),} \\
&= \frac{\Psi(n^{-1})}{P_{n-1}}.
\end{aligned}$$

Applying the Abel transformation to $I_4(n)$, we see that

$$\begin{aligned}\sum_{k=n}^{\infty} p_k \sin (n-k)t &= -\{p_{n+1}\sin t + p_{n+2}\sin 2t + \cdots\}\\ &= -\{(p_{n+1}-p_{n+2})\sin t\\ &\qquad + (p_{n+2}-p_{n+3})(\sin t + \sin 2t) + \cdots\}\\ &= -\sum_{k=n}^{\infty}(p_k - p_{k+1})\frac{\cos \frac{1}{2}t - \cos (n-k-\frac{1}{2})t}{2\sin \frac{1}{2}t};\end{aligned}$$

$$\begin{aligned}\sum_{k=0}^{n-1} P_k \sin (n-k)t &= P_0 \sin nt + \cdots + P_{n-1}\sin t\\ &= (P_{n-1}-P_{n-2})\sin t + (P_{n-2}-P_{n-3})(\sin t + \sin 2t)\\ &\qquad + \cdots + (P_1 - P_0)(\sin t + \cdots + \sin (n-1)t)\\ &\qquad + P_0(\sin t + \cdots + \sin nt)\\ &= \sum_{k=0}^{n-1} p_k \frac{\cos \frac{1}{2}t - \cos (n-k+\frac{1}{2})t}{2\sin \frac{1}{2}t}.\end{aligned}$$

Hence,

$$\begin{aligned}&\sum_{k=n}^{\infty} p_k \sin (n-k)t + \sum_{k=0}^{n-1}\frac{p_n}{P_n}P_k \sin (n-k)t\\ &\quad = \sum_{k=n}^{\infty}(p_k - p_{k+1})\frac{\cos (n-k-\frac{1}{2})t}{2\sin \frac{1}{2}t} - \frac{p_n}{P_n}\sum_{k=0}^{n-1} p_k \frac{\cos (n-k+\frac{1}{2})t}{2\sin \frac{1}{2}t}\\ &\qquad + \frac{\cos \frac{1}{2}t}{2\sin \frac{1}{2}t}p_n\left(\frac{P_{n-1}}{P_n} - 1\right).\end{aligned}$$

Therefore,

$$\begin{aligned}I_4(n) &\leqq \frac{1}{P_{n-1}}\left|\int_{1/n}^{\pi}\frac{\psi(t)}{2\sin \frac{1}{2}t}\sum_{k=n}^{\infty}(p_k - p_{k+1})\cos (n-k-\tfrac{1}{2})t\,dt\right|\\ &\qquad + \frac{p_n}{P_n P_{n-1}}\left|\int_{1/n}^{\pi}\frac{\psi(t)}{2\sin \frac{1}{2}t}\sum_{k=0}^{n-1} p_k \cos (n-k+\tfrac{1}{2})t\,dt\right|\\ &\qquad + \frac{Ap_n^2}{P_n P_{n-1}}\int_{1/n}^{\pi}\frac{|\psi(t)|}{2\tan \frac{1}{2}t}dt\\ &= I_{4,1}(n) + I_{4,2}(n) + I_{4,3}(n).\end{aligned}$$

Using the same method as in Lemma (5.16), we get

$$\sum_{n=1}^{\infty} I_{4,1}(n) + \sum_{n=1}^{\infty} I_{4,2}(n) \leqq A\left\{1 + \sum_{n=1}^{\infty}\frac{\Psi(n^{-1})}{P_{n-1}}\right\}.$$

Now

$$I_{4,3}(n) = A\frac{p_n^2}{P_nP_{n-1}}\int_{1/n}^{\pi}\frac{|\psi(t)|}{2\tan\frac{1}{2}t}dt \leqq A\frac{p_n}{P_nP_{n-1}}\int_{1/n}^{\pi}|\psi(t)|P(t^{-1})t^{-1}dt,$$

since $p_n \leqq p_0 = P_0 \leqq P(t^{-1})$, and, by the same argument as in Lemma (5.16),

$$\sum_{n=1}^{\infty} I_{4,3}(n) \leqq A\left\{1 + \sum_{n=1}^{\infty}\frac{\Psi(n^{-1})}{P_{n-1}}\right\};$$

hence, (5.19) is established.

(5.20) LEMMA. *If p_n is non-negative and non-increasing, and if we write $\gamma(t) = \sum_{k=0}^{\infty} p_k e^{ikt}$, then, for t in $[h, \pi]$,*

$$|\gamma(t+2h) - \gamma(t)| \leqq Aht^{-1}P(h^{-1}). \tag{5.21}$$

Proof. Let $\tau = [t^{-1}]$ and $\theta = [h^{-1}]$. It is clear that $\tau \leqq \theta$. Now

$$\begin{aligned}\gamma(t+2h) - \gamma(t) &= \sum_{k=0}^{\infty} p_k\{\exp(ik(t+2h)) - \exp(ikt)\}\\ &= \left(\sum_{k=0}^{\tau} + \sum_{k=\tau+1}^{\theta} + \sum_{k=\theta+1}^{\infty}\right) p_k\{\exp(ik(t+2h)) - \exp(ikt)\}\\ &= S_1 + S_2 + S_3,\end{aligned}$$

from which the term S_2 may be absent.

$$\begin{aligned}|S_1| &= \left|\sum_{k=0}^{\tau} p_k \exp(ik(t+h))2i\sin kh\right|\\ &\leqq 2\sum_{k=0}^{\tau} p_k kh, \qquad \text{since } kh \leqq ht^{-1} \leqq 1 \text{ for } 0 \leqq k \leqq \tau,\\ &\leqq 2ht^{-1}\sum_{k=0}^{\tau} p_k \leqq Aht^{-1}P(t^{-1}) \leqq Aht^{-1}P(h^{-1}).\end{aligned}$$

By (3.02), we have

$$\begin{aligned}|S_3| &\leqq Ap(h^{-1})\max_{\theta+1\leqq n\leqq\infty}\left|\sum_{k=0}^{n}(\exp(ik(t+2h)) - \exp(ikt))\right|\\ &= Ap(h^{-1})\max_{\theta+1\leqq n\leqq\infty}\left|\frac{1-\exp(i(n+1)(t+2h))}{1-\exp(i(t+2h))} - \frac{1-\exp(i(n+1)t)}{1-\exp(it)}\right|\\ &\leqq Ap(h^{-1})\left|\frac{1}{1-\exp(it)}\right| \leqq Ap(h^{-1})(\sin\tfrac{1}{2}t)^{-1} \leqq At^{-1}p(h^{-1})\\ &\leqq Aht^{-1}P(h^{-1}), \qquad \text{by Lemma (5.10).}\end{aligned}$$

Applying Abel's transformation to S_2, we get

$$\begin{aligned}
S_2 &= \sum_{k=\tau+1}^{\theta} p_k(\exp(ik(t+2h)) - \exp(ikt)) \\
&= \sum_{k=\tau+1}^{\theta-1} (p_k - p_{k+1}) \left\{ \frac{1-\exp(i(k+1)(t+2h))}{1-\exp(i(t+2h))} - \frac{1-\exp(i(k+1)t)}{1-\exp(it)} \right\} \\
&\qquad - p_{\tau+1} \left\{ \frac{1-\exp(i(\tau+1)(t+2h))}{1-\exp(i(t+2h))} - \frac{1-\exp(i(\tau+1)t)}{1-\exp(it)} \right\} \\
&\qquad + p_\theta \left\{ \frac{1-\exp(i(\theta+1)(t+2h))}{1-\exp(i(t+2h))} - \frac{1-\exp(i(\theta+1)t)}{1-\exp(it)} \right\} \\
&= S_{2,1} + S_{2,2} + S_{2,3}.
\end{aligned}$$

$$|S_{2,3}| \leqq Ap(h^{-1}) \left| \frac{1}{1-\exp(it)} \right| \leqq At^{-1}p(h^{-1}) \leqq Aht^{-1}P(h^{-1}).$$

$$\begin{aligned}
S_{2,2} &= p_{\tau+1} \left\{ \frac{\exp(i(\tau+1)(t+2h))}{1-\exp(i(t+2h))} - \frac{\exp(i(\tau+1)t)}{1-\exp(it)} \right. \\
&\qquad \left. - \frac{(\exp(it))(\exp(i2h)-1)}{(1-\exp(i(t+2h)))(1-\exp(it))} \right\}.
\end{aligned} \tag{5.22}$$

Let us consider the last term in the braces.

$$\begin{aligned}
&\left| \frac{(\exp(it))(\exp(i2h)-1)}{(1-\exp(i(t+2h)))(1-\exp(it))} \right| \\
&\qquad = \left| \frac{2i(\exp(it))(\exp(ih))\sin h}{-2(\exp(\tfrac{1}{2}i(t+2h)))(\exp(\tfrac{1}{2}it))\sin(\tfrac{1}{2}(t+2h))2\sin\tfrac{1}{2}t} \right| \qquad (5.23) \\
&\qquad = \frac{1}{2} \frac{\sin h}{\sin(\tfrac{1}{2}(t+2h))\sin\tfrac{1}{2}t} \leqq Aht^{-2}.
\end{aligned}$$

Next, we consider the remaining terms in the braces on the right of (5.22).

$$\begin{aligned}
\frac{\exp(i(\tau+1)t)}{1-\exp(it)} &= -\frac{\exp(i(\tau+1)t)}{2i(\exp(\tfrac{1}{2}it))\sin\tfrac{1}{2}t} = \frac{i}{2} \frac{\exp(i(\tau+\tfrac{1}{2})t)}{\sin\tfrac{1}{2}t} \\
&= \frac{i}{2} \frac{\cos(\tau+\tfrac{1}{2})t + i\sin(\tau+\tfrac{1}{2})t}{\sin\tfrac{1}{2}t} \\
&= f_\tau(t) + ig_\tau(t),
\end{aligned}$$

where

$$f_\tau(t) = -\frac{\sin(\tau+\tfrac{1}{2})t}{2\sin\tfrac{1}{2}t}, \qquad g_\tau(t) = \frac{\cos(\tau+\tfrac{1}{2})t}{2\sin\tfrac{1}{2}t}.$$

Hence,

$$\begin{aligned}
&\frac{\exp(i(\tau+1)(t+2h))}{1-\exp(i(t+2h))} - \frac{\exp(i(\tau+1)t)}{1-\exp(it)} \qquad (5.24) \\
&\qquad = f_\tau(t+2h) - f_\tau(t) + i\{g_\tau(t+2h) - g_\tau(t)\}.
\end{aligned}$$

By the mean value theorem,

$$f_\tau(t+2h) - f_\tau(t) = 2hf'_\tau(t+2\theta_1 h), \qquad 0 \leqq \theta_1 \leqq 1,$$

$$= h \frac{\left\{\frac{1}{2}\sin((\tau+\frac{1}{2})(t+2\theta_1 h))\cos\left(\frac{t+2\theta_1 h}{2}\right) - \sin\left(\frac{t+2\theta_1 h}{2}\right) \times ((\tau+\frac{1}{2})\cos((\tau+\frac{1}{2})(t+2\theta_1 h)))\right\}}{\sin^2\frac{t+2\theta_1 h}{2}}.$$

Therefore, $|f_\tau(t+2h) - f_\tau(t)| \leqq Aht^{-2}$, and similarly $|g_\tau(t+2h) - g_\tau(t)| \leqq Aht^{-2}$. Hence

$$|S_{2,2}| \leqq Aht^{-2}p(t^{-1}) \leqq Aht^{-1}P(t^{-1}) \leqq Aht^{-1}P(h^{-1}).$$

$$S_{2,1} = \sum_{k=\tau+1}^{\theta-1}(p_k - p_{k+1})\left\{\frac{1-\exp(i(k+1)(t+2h))}{1-\exp(i(t+2h))} - \frac{1-\exp(i(k+1)t)}{1-\exp(it)}\right\}$$

$$= \sum_{k=\tau+1}^{\theta-1}(p_k - p_{k+1})\left\{-\frac{\exp(i(k+1)(t+2h))}{1-\exp(i(t+2h))} + \frac{\exp(i(k+1)t)}{1-\exp(it)} + \frac{(\exp(it))((\exp(i2h))-1)}{(1-\exp(i(t+2h)))(1-\exp(it))}\right\}.$$

We have already shown in (5.23) that

$$\left|\frac{(\exp(it))((\exp(i2h))-1)}{(1-\exp(i(t+2h)))(1-\exp(it))}\right| \leqq Aht^{-2},$$

and, since $\sum_{\tau+1}^{\theta-1}(p_k - p_{k+1}) = p_{\tau+1} - p_\theta \leqq p_{\tau+1} \leqq p(t^{-1})$, we have

$$\left|\sum_{k=\tau+1}^{\theta-1}(p_k - p_{k+1})\frac{(\exp(it))((\exp(i2h))-1)}{(1-\exp(i(t+2h)))(1-\exp(it))}\right| \leqq Aht^{-2}p(t^{-1}) \leqq Aht^{-1}P(h^{-1}).$$

As in (5.24), we have

$$\frac{\exp(i(k+1)(t+2h))}{1-\exp(i(t+2h))} - \frac{\exp(i(k+1)t)}{1-\exp(it)} = f_k(t+2h) - f_k(t) + i\{g_k(t+2h) - g_k(t)\},$$

so that, by the mean value theorem, we have

$$|f_k(t+2h) - f_k(t)| = |2hf'_k(t+2\theta_1 h)|, \qquad 0 \leqq \theta_1 \leqq 1,$$

$$\left| \tfrac{1}{2} \sin\left((k+\tfrac{1}{2})(t+2\theta_1 h)\right) \cos\left(\frac{t+2\theta_1 h}{2}\right) - \sin\left(\frac{t+2\theta_1 h}{2}\right) \right.$$

$$\left. \times (k+\tfrac{1}{2}) \cos\left((k+\tfrac{1}{2})(t+2\theta_1 h)\right) \right|$$

$$= h \frac{}{\sin^2 \dfrac{t+2\theta_1 h}{2}}$$

$$\leqq Aht^{-2} + Aht^{-1}k,$$

and, in a similar manner, $|\, g_k(t+2h) - g_k(t) \,| \leqq Aht^{-2} + Aht^{-1}k$. Therefore,

$$\left| \sum_{k=\tau+1}^{\theta-1} (p_k - p_{k+1}) \left\{ - \frac{\exp\left(i(k+1)(t+2h)\right)}{1-\exp\left(i(t+2h)\right)} + \frac{\exp\left(i(k+1)t\right)}{1-\exp\left(it\right)} \right\} \right|$$

$$\leqq A \sum_{k=\tau+1}^{\theta-1} (p_k - p_{k+1})(ht^{-2} + ht^{-1}k)$$

$$\leqq Aht^{-2}p(t^{-1}) + Aht^{-1} \sum_{k=\tau+1}^{\theta-1} (p_k - p_{k+1})k$$

$$\leqq Aht^{-1}P(t^{-1}) + Aht^{-1}\{(\tau+1)p_{\tau+1} + P_\theta\}$$

$$\leqq Aht^{-1}P(h^{-1}) + Aht^{-2}p(t^{-1}) + Aht^{-1}P(h^{-1})$$

$$\leqq Aht^{-1}P(h^{-1}).$$

Hence, $|\, S_{2,1} \,| \leqq Aht^{-1}P(h^{-1})$. Therefore,

$$|\, \gamma(t+2h) - \gamma(t) \,| \leqq |\, S_1 \,| + |\, S_3 \,| + |\, S_{2,1} \,| + |\, S_{2,2} \,| + |\, S_{2,3} \,| \leqq Aht^{-1}P(h^{-1}).$$

(5.25) DEFINITION. $f(x)$ is said to belong to the class Lip α for $a \leqq x \leqq b$ if $|\, f(x+h) - f(x) \,| \leqq A\,|\, h \,|^\alpha$ for $a \leqq x \leqq b$.

We will now prove the following theorem.

(5.26) THEOREM. *Let N_p be a Nörlund transformation with p_n non-negative and non-increasing, $\lim_{n\to\infty} p_n = 0$, $|\, \Delta p_n \,|$ non-increasing and satisfying the conditions* (1) $\sum_{k=1}^{\infty} P_k^2 k^{-2} \leqq A$ *and* (2) $\sum_{k=1}^{\infty} P_k^{-1} k^{-\alpha-1} \leqq A$. *If $f(x)$ belongs to class Lip α, $0 < \alpha \leqq 1$, then $S(f)$ and $\tilde{S}(f)$ are both summable $|\, N_p \,|$.*

The following lemmas are pertinent to the proof of this theorem.

(5.27) LEMMA. *If $f(x)$ belongs to Lip α, then $\phi(t)$ and $\psi(t)$ also belong to Lip α.*

Proof.

$$|\, \phi(t+h) - \phi(t) \,| = |\, f(x+t+h) + f(x-t-h) - f(x+t) - f(x-t) \,|$$

$$\leqq |\, f(x+t+h) - f(x+t) \,| + |\, f(x-t-h) - f(x-t) \,|$$

$$\leqq A\,|\, h \,|^\alpha.$$

Similarly, $|\psi(t+h)-\psi(t)| \leqq A|h|^{\alpha}$.

(5.28) LEMMA. *If $f(x)$ belongs to Lip α and* (5.26), (1) *is satisfied, then $\phi(t)P(t^{-1})$ and $\psi(t)P(t^{-1})$ belong to L^1.*

Proof. By Lemma (5.27), $\phi(t)$ and $\psi(t)$ belong to Lip α; they are therefore uniformly continuous and hence bounded on $[0, \pi]$. Consequently

$$\begin{aligned}\int_0^{\pi} |\phi(t)|\,P(t^{-1})\,dt &\leqq A\int_0^{\pi} P(t^{-1})\,dt\\ &= A\int_{1/\pi}^{\infty} t^{-2}P(t)\,dt\\ &= \int_{1/\pi}^{1} t^{-2}P(t)\,dt + \int_1^{\infty} t^{-2}P(t)\,dt\\ &\leqq A + \sum_{k=1}^{\infty} P_k k^{-2} \leqq A + A\sum_{k=1}^{\infty} P_k^2 k^{-2} \leqq A.\end{aligned}$$

A similar result holds for $\psi(t)$.

(5.29) LEMMA. *If $f(x)$ belongs to Lip α, $\alpha > 0$, and conditions* (5.26), (1) *and* (5.26), (2) *are satisfied, then*

$$\sum_{n=1}^{\infty} \Phi(n^{-1})P_{n-1}^{-1} \leqq A, \qquad \sum_{n=1}^{\infty} \Psi(n^{-1})P_{n-1}^{-1} \leqq A.$$

Proof. Clearly $|\phi(t)| \leqq At^{\alpha}$. Therefore

$$\begin{aligned}\Phi(n^{-1}) &\leqq A\int_0^{1/n} t^{\alpha}P(t^{-1})\,dt\\ &= A\int_n^{\infty} t^{-\alpha-2}P(t)\,dt\\ &\leqq A\sum_{k=n}^{\infty} P_k k^{-\alpha-2}\\ &\leqq A\left\{\sum_{k=n}^{\infty} P_k^2 k^{-2}\right\}^{\frac{1}{2}}\left\{\sum_{k=n}^{\infty} k^{-2-2\alpha}\right\}^{\frac{1}{2}} && \text{(Schwarz's inequality)}\\ &\leqq A\left\{\sum_{k=n}^{\infty} k^{-2-2\alpha}\right\}^{\frac{1}{2}}, && \text{by (5.26), (1),}\\ &\leqq An^{-\alpha-\frac{1}{2}}.\end{aligned}$$

Hence

$$\sum_{n=1}^{\infty} \Phi(n^{-1})P_{n-1}^{-1} \leqq A\sum_{n=1}^{\infty} P_n^{-1}n^{-\alpha-\frac{1}{2}} \leqq A, \qquad \text{by (5.26), (2).}$$

The second inequality follows in a similar manner.

(5.30) LEMMA. *If p_n is non-negative and non-increasing and $\sum_{k=1}^{\infty} P_k^2 k^{-2} \leqq A$, then $P_n^2 n^{-1} \leqq A$.*

Proof. Let $\gamma_k = P_k^2 k^{-1}$. Then

$$\begin{aligned}\gamma_{k+1} - \gamma_k &= P_{k+1}^2 (k+1)^{-1} - P_k^2 k^{-1} \\ &= k^{-1}(k+1)^{-1}\{kP_{k+1}^2 - (k+1)P_k^2\} \\ &= k^{-1}(k+1)^{-1}\{2kp_{k+1}P_k + kp_{k+1}^2 - P_k^2\} \\ &\leqq AP_k^2 k^{-2}.\end{aligned}$$

Therefore,

$$\gamma_n - \gamma_1 = \sum_{k=1}^{n-1} (\gamma_{k+1} - \gamma_k) \leqq A \sum_{k=1}^{n} P_k^2 k^{-2} \leqq A.$$

Hence $\gamma_n \leqq A$.

(5.31) LEMMA. *If p_n is non-negative and non-increasing, then*

$$\sum_{\nu=1}^{\infty} P_{2^{\nu-1}}^{-1} 2^{-\nu(\alpha-\frac{1}{2})} \leqq A \sum_{n=1}^{\infty} P_n^{-1} n^{-\alpha-\frac{1}{2}}.$$

Proof. First we observe that

$$\begin{aligned}P_{2^\nu} P_{2^{\nu-1}}^{-1} &= (P_{2^{\nu-1}} + p_{2^{\nu-1}+1} + \cdots + p_{2^\nu}) P_{2^{\nu-1}}^{-1} \\ &\leqq 1 + 2^{\nu-1} p_{2^{\nu-1}} P_{2^{\nu-1}}^{-1} \leqq A.\end{aligned} \tag{5.32}$$

Therefore,

$$\begin{aligned}\sum_{\nu=1}^{\infty} P_{2^{\nu-1}}^{-1} 2^{-\nu(\alpha-\frac{1}{2})} &\leqq A \sum_{\nu=1}^{\infty} P_{2^{\nu-1}}^{-1} \sum_{k=2^{\nu-1}+1}^{2^\nu} k^{-\alpha-\frac{1}{2}} \\ &\leqq A \sum_{\nu=1}^{\infty} \sum_{k=2^{\nu-1}+1}^{2^\nu} P_k^{-1} k^{-\alpha-\frac{1}{2}} \\ &\leqq A \sum_{n=1}^{\infty} P_n^{-1} n^{-\alpha-\frac{1}{2}}.\end{aligned}$$

Proof of (5.26). In view of our hypotheses and Lemma (5.28), it is clear that the conditions of Lemmas (5.16) and (5.18) are satisfied; therefore, the inequalities (5.17) and (5.19) are true. Hence, due to Lemma (5.29), it suffices to show that

$$\sum_{n=1}^{\infty} \frac{|\alpha_n| + |\beta_n|}{P_{n-1}} \leqq A, \qquad \sum_{n=1}^{\infty} \frac{|\tilde{\alpha}_n| + |\tilde{\beta}_n|}{P_{n-1}} \leqq A.$$

Now

$$\int_0^\pi \alpha^2(t)\,dt \leqq A\int_0^\pi P^2(t^{-1})\,dt, \qquad \text{by Lemma (5.11),}$$

$$= A\int_{1/\pi}^\infty P^2(t)t^{-2}\,dt \tag{5.33}$$

$$\leqq A + A\sum_{k=1}^\infty P_k^2 k^{-2} \leqq A, \qquad \text{by (5.26), (1).}$$

Hence, since $\phi(t)$ is bounded, α_n is the Fourier coefficient of an even function belonging to L^2. Similarly, $\tilde{\alpha}_n$ is the Fourier coefficient of an even function belonging to L^2, and β_n and $\tilde{\beta}_n$ are Fourier coefficients of odd functions belonging to L^2. The Fourier series of $\phi(t+h)\alpha(t+h) - \phi(t-h)\alpha(t-h)$ is

$$-\frac{4}{\pi}\sum_{n=1}^\infty \alpha_n \sin nt \sin nh.$$

On applying Parseval's relation, we get

$$\sum_{n=1}^\infty \alpha_n^2 \sin^2 nh \leqq A\int_0^\pi \{\phi(t+h)\alpha(t+h) - \phi(t-h)\alpha(t-h)\}^2\,dt$$

$$\leqq A\left(\int_0^\pi \alpha^2(t+h)\{\phi(t+h) - \phi(t-h)\}^2\,dt + \int_0^\pi \phi^2(t-h)\{\alpha(t+h) - \alpha(t-h)\}^2\,dt\right)$$

$$= J_1(h) + J_2(h).$$

Taking h to be positive, it is clear from (5.27) and (5.33) that $J_1(h) \leqq Ah^{2\alpha}$. Now

$$J_2(h) = A\int_0^\pi \phi^2(t-h)\{\alpha(t+h) - \alpha(t-h)\}^2\,dt$$

$$= A\int_{-h}^{\pi-h} \phi^2(t)\{\alpha(t+2h) - \alpha(t)\}^2\,dt$$

$$\leqq A\int_{-h}^{h} \phi^2(t)\alpha^2(t+2h)\,dt + A\int_{-h}^{h} \phi^2(t)\alpha^2(t)\,dt + A\int_{h}^{\pi-h} \phi^2(t)\{\alpha(t+2h) - \alpha(t)\}^2\,dt$$

$$= J_{2,1}(h) + J_{2,2}(h) + J_{2,3}(h).$$

Consider each of these integrals separately. First,

$$J_{2,1}(h) \leqq A\int_{-h}^{h} t^{2\alpha} P^2\left(\frac{1}{t+2h}\right)dt$$

$$\leqq Ah^{2\alpha+1} P^2(h^{-1})$$

$$\leqq Ah^{2\alpha}, \qquad \text{by Lemma (5.30).}$$

Next,

$$\begin{aligned} J_{2,2}(h) &\leqq A\int_0^h t^{2\alpha}P^2(t^{-1})\,dt \\ &= A\int_{h^{-1}}^{\infty} t^{-2\alpha-2}P^2(t)\,dt \\ &\leqq A\sum_{k=[h^{-1}]}^{\infty} P_k^2 k^{-2\alpha-2} \\ &\leqq Ah^{2\alpha}\sum_{k=[h^{-1}]}^{\infty} P_k^2 k^{-2} \\ &\leqq Ah^{2\alpha}, \qquad \text{by (5.26), (1).} \end{aligned}$$

Finally,

$$\begin{aligned} J_{2,3}(h) &\leqq Ah^2P^2(h^{-1})\int_h^{\pi} t^{2\alpha-2}, \qquad \text{by Lemma (5.20),} \\ &\leqq \begin{cases} Ah^2P^2(h^{-1})\dfrac{\pi^{2\alpha-1}-h^{2\alpha-1}}{2\alpha-1}, & \alpha \neq \frac{1}{2}, \\ Ah^2P^2(h^{-1})(\log\pi - \log h), & \alpha = \frac{1}{2}, \end{cases} \\ &\leqq \begin{cases} Ah^{2\alpha+1}P^2(h^{-1}), & \alpha < \frac{1}{2}, \\ Ah^2P^2(h^{-1}), & \alpha > \frac{1}{2}, \\ Ah^2P^2(h^{-1})\log h^{-1}, & \alpha = \frac{1}{2}, \end{cases} \\ &\leqq \begin{cases} Ah^{2\alpha}, & \alpha < \frac{1}{2}, \text{ by Lemma (5.30),} \\ Ah^2P^2(h^{-1}), & \alpha > \frac{1}{2}, \\ Ah^2P^2(h^{-1})\log h^{-1}, & \alpha = \frac{1}{2}. \end{cases} \end{aligned}$$

Therefore,

$$\sum_{n=1}^{\infty}\alpha_n^2\sin^2 nh \leqq Ah^{2\alpha} + Ah^2P^2(h^{-1})\log h^{-1}.$$

Now let us write $h = \pi/2N$; then the above inequality becomes

$$\begin{aligned} \sum_{n=1}^{\infty}\alpha_n^2\sin^2(n\pi/2N) &\leqq A\{N^{-2\alpha} + N^{-2}P^2(2N/\pi)\log(2N/\pi)\} \\ &\leqq A\{N^{-2\alpha} + N^{-2}P^2(N)\log N\}. \end{aligned}$$

Taking $N = 2^{\nu}$, we get

$$\begin{aligned} \sum_{n=2^{\nu-1}+1}^{2^{\nu}}\alpha_n^2 &\leqq 2\sum_{n=2^{\nu-1}+1}^{2^{\nu}}\alpha_n^2\sin^2(n\pi/2^{\nu+1}) \\ &\leqq 2\sum_{n=1}^{\infty}\alpha_n^2\sin^2(n\pi/2^{\nu+1}) \\ &\leqq A\{2^{-2\nu\alpha} + \nu 2^{-2\nu}P_{2^{\nu}}^2\}. \end{aligned}$$

Therefore, applying Schwarz's inequality, we get

$$\begin{aligned}\sum_{n=2^{v-1}+1}^{2^v} |\alpha_n| P_n^{-1} &\leqq \left\{\sum_{n=2^{v-1}+1}^{2^v} \alpha_n^2\right\}^{\frac{1}{2}} \left\{\sum_{n=2^{v-1}+1}^{2^v} P_n^{-2}\right\}^{\frac{1}{2}} \\ &\leqq A\{2^{-v\alpha} + v^{\frac{1}{2}} 2^{-v} P_{2^v}\} \cdot 2^{(v-1)/2} P_{2^{v-1}}^{-1} \\ &= A\{2^{-v(\alpha-\frac{1}{2})} P_{2^{v-1}}^{-1} + v^{\frac{1}{2}} 2^{-\frac{1}{2}(v+1)} P_{2^v} P_{2^{v-1}}^{-1}\} \\ &\leqq A\{2^{-v(\alpha-\frac{1}{2})} P_{2^{v-1}}^{-1} + v^{\frac{1}{2}} 2^{-\frac{1}{2}(v+1)}\}, \qquad \text{by (5.32).}\end{aligned}$$

Hence,

$$\begin{aligned}\sum_{n=1}^{\infty} |\alpha_n| P_n^{-1} &= A \sum_{v=1}^{\infty} \sum_{n=2^{v-1}+1}^{2^v} |\alpha_n| P_n^{-1} \\ &\leqq A \sum_{v=1}^{\infty} \{P_{2^{v-1}}^{-1} 2^{-v(\alpha-\frac{1}{2})} + v^{\frac{1}{2}} 2^{-\frac{1}{2}(v+1)}\} \\ &\leqq A \sum_{n=1}^{\infty} \{P_n^{-1} n^{-(\alpha+\frac{1}{2})} + n^{\frac{1}{2}} 2^{-\frac{1}{2}(n+1)}\}, \qquad \text{by Lemma (5.31),} \\ &\leqq A + A \sum_{n=1}^{\infty} P_n^{-1} n^{-(\alpha+\frac{1}{2})} \\ &\leqq A, \qquad \text{by (5.26), (2).}\end{aligned}$$

It can be shown in a similar manner that $\sum_{n=1}^{\infty} |\beta_n| P_n^{-1}$, $\sum_{n=1}^{\infty} |\tilde{\alpha}_n| P_n^{-1}$ and $\sum_{n=1}^{\infty} |\tilde{\beta}_n| P_n^{-1}$ are each bounded, and so our theorem is proved.

(5.34) COROLLARY (Bernstein's theorem). *If $f(x)$ belongs to Lip α, $\frac{1}{2} < \alpha \leqq 1$, then $S(f)$ and $\tilde{S}(f)$ converge absolutely.*

Proof. Absolute convergence is summability $|C, 0|$, or summability $|N_p|$, where $p_0 = 1$, $p_n = 0$, $n \neq 1$. Clearly p_n is non-negative and non-increasing, $\lim_{n\to\infty} p_n = 0$ and $|\Delta p_n|$ is non-increasing. Thus it suffices to show that (5.26), (1) and (2) are satisfied. Now

$$(1) \quad \sum_{k=1}^{\infty} P_k^2 k^{-2} = \sum_{k=1}^{\infty} k^{-2} \leqq A,$$

$$(2) \quad \sum_{k=1}^{\infty} P_k^{-1} k^{-\alpha-\frac{1}{2}} = \sum_{k=1}^{\infty} k^{-\alpha-\frac{1}{2}} \leqq A,$$

since $\alpha + \frac{1}{2} > 1$.

(5.35) COROLLARY (Hyslop's theorem). *If $f(x)$ belongs to Lip α, $0 < \alpha \leqq \frac{1}{2}$, and if $\beta + \alpha > \frac{1}{2}$, $\beta > 0$, then $S(f)$ and $\tilde{S}(f)$ are summable $|C, \beta|$.*

Proof. By Theorem (2.20), $|C, \beta| \subset |C, \gamma|$ for $\beta < \gamma$. Hence we can assume $0 < \beta < \frac{1}{2}$, and we only need to show that the conditions of Theorem

(5.26) are satisfied in this case. It is clear that p_n is non-negative and non-increasing, $\lim_{n\to\infty} p_n = 0$ and $|\Delta p_n|$ is non-increasing. Now

$$P_n = \frac{(\beta+1)(\beta+2)\cdots(\beta+n)}{n!} \simeq n^{\beta}.$$

Therefore,

(1) $\sum_{k=1}^{\infty} P_k^2 k^{-2} \leqq A \sum_{k=1}^{\infty} k^{2(\beta-1)} \leqq A$, since $2(\beta-1) < -1$;

(2) $\sum_{k=1}^{\infty} P_k^{-1} k^{-(\alpha+\frac{1}{2})} \leqq A \sum_{k=1}^{\infty} k^{-(\alpha+\beta+\frac{1}{2})} \leqq A$, since $\alpha+\beta+\frac{1}{2} > 1$.

(5.36) COROLLARY. *Let the Nörlund transformation N_p be defined by*

$$p_n = \frac{(n+c)^{\beta-1}\log^{1+\epsilon}(n+c)}{c^{\beta-1}\log^{1+\epsilon} c},$$

where $0 < \beta < \frac{1}{2}$, $\epsilon > 0$ and $\log c \geqq \frac{8}{3}(1+\epsilon)$. Let $f(x)$ belong to class Lip α, where $\alpha + \beta = \frac{1}{2}$. Then $S(f)$ and $\tilde{S}(f)$ are both summable $|N_p|$.

Proof. Obviously $p_n \geqq 0$ and $p_0 = 1$. We will write

$$p(x) = \frac{(x+c)^{\beta-1}\log^{1+\epsilon}(x+c)}{c^{\beta-1}\log^{1+\epsilon} c}.$$

Then

$$p'(x) = \frac{(x+c)^{\beta-2}\log^{\epsilon}(x+c)}{c^{\beta-1}\log^{1+\epsilon} c}\{1+\epsilon+(\beta-1)\log(x+c)\} \leqq 0$$

for all $x \geqq 0$ when $\log c > 2(1+\epsilon)$. Therefore, $p(x)$, and consequently $p_n = p(n)$, is non-increasing. Also,

$$p''(x) = \frac{(x+c)^{\beta-3}\log^{\epsilon}(x+c)}{c^{\beta-1}\log^{1+\epsilon} c}\left\{(1+\epsilon)(2\beta-3) + \frac{\epsilon(\epsilon+1)}{\log(x+c)} + (\beta-1)(\beta-2)\log(x+c)\right\}$$

$$\geqq 0$$

for all $x \geqq 0$ when $\log c \geqq \frac{8}{3}(1+\epsilon)$. Therefore $|p'(x)|$, and hence also $|\Delta p_n|$, is non-increasing. Finally, $P_n \simeq n^{\beta}\log^{1+\epsilon} n$. Therefore,

(1) $\sum_{k=1}^{\infty} P_k^2 k^{-2} \leqq A \sum_{k=1}^{\infty} \frac{\log^{2+2\epsilon} k}{k^{2(1-\beta)}} \leqq A$, since $1-\beta > \frac{1}{2}$.

(2) $\sum_{k=1}^{\infty} P_k^{-1} k^{-\alpha-\frac{1}{2}} \leqq A \sum_{k=1}^{\infty} (\log^{1+\epsilon} k \cdot k^{\alpha+\beta+\frac{1}{2}})^{-1}$

$$= A \sum_{k=1}^{\infty} (k\log^{1+\epsilon} k)^{-1}$$

$$\leqq A.$$

(5.37) COROLLARY. *Let the Nörlund transformation N_p be defined by*

$$p_n = \frac{c \log^{\epsilon}(n+c)}{(n+c)\log^{\epsilon} c},$$

where $\log c \geqq \frac{3}{2}\epsilon > 0$. *Let $f(x)$ belong to class Lip $\frac{1}{2}$. Then $S(f)$ and $\tilde{S}(f)$ are summable $|N_p|$.*

Proof. Clearly $p_n \geqq 0$ and $p_0 = 1$. If we write

$$p(x) = \frac{c \log^{\epsilon}(x+c)}{(x+c)\log^{\epsilon} c},$$

then

$$p'(x) = \frac{c \log^{\epsilon}(x+c)}{(x+c)^2 \log^{\epsilon} c}\left\{\frac{\epsilon}{\log(x+c)} - 1\right\} \leqq 0$$

for $x \geqq 0$ and $\log c \geqq \epsilon$. Therefore, p_n is non-increasing. Also

$$\begin{aligned} p''(x) &= \frac{c \log^{\epsilon}(x+c)}{(x+c)^3 \log^{\epsilon} c}\left\{\frac{\epsilon(\epsilon-1)}{\log^2(x+c)} - \frac{2\epsilon}{\log(x+c)} - \frac{\epsilon}{\log(x+c)} + 2\right\} \\ &\geqq \frac{c \log^{\epsilon}(x+c)}{(x+c)^3 \log^{\epsilon} c}\left\{2 - \frac{3\epsilon}{\log(x+c)}\right\} \\ &\geqq 0 \end{aligned}$$

for $x \geqq 0$ and $\log c \geqq \frac{3}{2}\epsilon$. Therefore, $|\Delta p_n|$ is non-increasing. Finally, $P_n \simeq \log^{1+\epsilon} n$. Therefore,

(1) $$\sum_{k=1}^{\infty} P_k^2 k^{-2} \leqq A \sum_{k=1}^{\infty} \frac{\log^{2+2\epsilon} k}{k^2} \leqq A.$$

(2) $$\sum_{k=1}^{\infty} P_k^{-1} k^{-1} \leqq A \sum_{k=1}^{\infty} (k \log^{1+\epsilon} k)^{-1} \leqq A.$$

(5.38) DEFINITION. $f(x)$ is said to belong to class Lip (α, q) for $a \leqq x \leqq b$ if

$$\left\{\int_a^b |f(x+h) - f(x)|^q \, dx\right\}^{1/q} \leqq A |h|^{\alpha}.$$

(5.39) THEOREM. *Let N_p be a Nörlund transformation with p_n non-negative and non-increasing,* $\lim_{n\to\infty} p_n = 0$, *$|\Delta p_n|$ non-increasing, and satisfying the conditions*

(1) $$\sum_{k=1}^{\infty} P_k^{2q/(q-2)} k^{-2} \leqq A, \qquad 2 < q < \infty,$$

(2) $$\sum_{k=1}^{\infty} P_k^{-1} k^{-\alpha-1} \leqq A.$$

If $f(x)$ belongs to class Lip (α, q) on $[0, \pi]$, $0 < \alpha \leqq 1$, $\alpha q > 1$, then $S(f)$ and $\tilde{S}(f)$ are both summable $|N_p|$.

The following lemmas are pertinent to the proof of this theorem.

(5.40) LEMMA. *If $f(x)$ belongs to Lip (α, q) on $[0, \pi]$, then $\phi(t)$ and $\psi(t)$ belong to Lip (α, q) on $[0, \pi]$.*

Proof. Clearly

$$|\phi(t+h) - \phi(t)| \leqq |f(x+t+h) - f(x+t)| + |f(x-t-h) - f(x-t)|.$$

Hence, by Minkowski's inequality,

$$\left\{\int_0^\pi |\phi(t+h) - \phi(t)|^q \omega_t\right\}^{\frac{1}{q}} \leqq \left\{\int_0^\pi |f(x+t+h) - f(x+t)|^q \, dx\right\}^{\frac{1}{q}} + \left\{\int_0^\pi |f(x-t-h) - f(x-t)|^q \, dx\right\}^{\frac{1}{q}} \leqq A|h|^\alpha.$$

A similar proof holds for $\psi(t)$.

(5.41) LEMMA. *If $f(x)$ belongs to Lip (α, q), $q \geqq 1$, $0 < \alpha \leqq 1$, $\alpha q > 1$, then $f(x)$ is equivalent to a function of class Lip $(\alpha - 1/q)$.*

Proof. This result was obtained by Hardy and Littlewood [4].

(5.42) LEMMA. *If $f(x)$ belongs to Lip (α, q), $q \geqq 1$, $0 < \alpha \leqq 1$, $\alpha q > 1$, and condition* (5.39), (1) *is satisfied, then $\phi(t)P(t^{-1})$ and $\psi(t)P(t^{-1})$ belong to class L^1.*

Proof. Due to Lemmas (5.40) and (5.41), $\phi(t)$ and $\psi(t)$ belong to class Lip $(\alpha - 1/q)$, and are therefore uniformly continuous and hence bounded on $[0, \pi]$. Therefore

$$\begin{aligned}\int_0^\pi |\phi(t)| P(t^{-1}) \, dt &\leqq A \int_0^\pi P(t^{-1}) \, dt \\ &= A \int_{1/\pi}^\infty t^{-2} P(t) \, dt \\ &\leqq A + A \sum_{k=1}^\infty P_k k^{-2} \\ &\leqq A + A \sum_{k=1}^\infty P_k^{2q/(q-2)} k^{-2} \leqq A.\end{aligned}$$

The result for $\psi(t)P(t^{-1})$ follows in the same manner.

(5.43) LEMMA. *If $f(x)$ belongs to Lip (α, q), $q \geqq 1$, $0 < \alpha \leqq 1$, $\alpha q > 1$, and conditions* (5.39), (1) *and* (5.39), (2) *are satisfied, then*

$$\sum_{n=1}^\infty \Phi(n^{-1}) P_{n-1}^{-1} \leqq A, \qquad \sum_{n=1}^\infty \Psi(n^{-1}) P_{n-1}^{-1} \leqq A.$$

Proof. By Lemma (5.40), $\phi(t)$ belongs to class Lip (α, q), and, by Lemma (5.41), $\phi(t)$ is equivalent to a function of class Lip $(\alpha - 1/q)$. Hence,

$$\begin{aligned}\Phi(n^{-1}) &\leqq A\int_0^{1/n} t^{\alpha-1/q}\,P(t^{-1})\,dt\\ &= A\int_n^{\infty} t^{-\alpha+1/q-2}\,P(t)\,dt\\ &\leqq \sum_{k=n}^{\infty} P_k k^{-2-\alpha+1/q}\\ &\leqq A\left\{\sum_{k=n}^{\infty} P_k^{2q/(q-2)}\,k^{-2}\right\}^{(q-2)/2q}\left\{\sum_{k=n}^{\infty} k^{-(1+\alpha+1/q)(2q/(q+2))}\right\}^{(q+2)/2q}\end{aligned}$$

by Hölder's inequality. However,

$$\begin{aligned}&\left\{\sum_{k=n}^{\infty} k^{-(1+\alpha+1/q)(2q/(q+2))}\right\}^{(q+2)/2q}\\ &\qquad\leqq A\,\{n^{-(2q+2q\alpha+2-q-2)/(q+2)}\}^{(q+2)/2q}\\ &\qquad\leqq An^{-\alpha-\frac{1}{2}}.\end{aligned}$$

Therefore, $\Phi(n^{-1}) \leqq An^{-\alpha-\frac{1}{2}}$ and consequently

$$\sum_{n=1}^{\infty}\Phi(n^{-1})\,P_{n-1}^{-1} \leqq A\sum_{n=1}^{\infty} P_{n-1}^{-1}\,n^{-\alpha-\frac{1}{2}} \leqq A.$$

The second inequality follows in a similar manner.

(5.44) LEMMA. *If* $\sum_{n=1}^{\infty} P_n^{2q/(q+2)}/n^2 \leqq A$ *for* $q > 2$, *then* $P_n^2/n^{1-2/q} \leqq A$.

Proof. First, we observe that for a given $r > 0$

$$\begin{aligned}\frac{P_n^r}{n} - \frac{P_{n+1}^r}{n+1} &= \frac{(n+1)P_n^r - nP_{n+1}^r}{n(n+1)}\\ &= \frac{P_n^r}{n(n+1)} + \frac{n(P_n^r - P_{n+1}^r)}{n(n+1)}.\end{aligned}$$

Now,

$$\begin{aligned}P_n^r - P_{n+1}^r &= P_n^r - (P_n + p_{n+1})^r\\ &= P_n^r - P_n^r\left(1 + \frac{p_{n+1}}{P_n}\right)^r\\ &= P_n^r - P_n^r\left(1 + r\frac{p_{n+1}}{P_n} + \frac{r(r-1)}{2!}\frac{p_{n+1}^2}{P_n^2} + \cdots\right)\\ &= -P_n^r\left(r\frac{p_{n+1}}{P_n} + \frac{r(r-1)}{2!}\frac{p_{n+1}^2}{P_n^2} + \cdots\right).\end{aligned}$$

However, $p_{n+1}/P_n \leqq p_n/P_n \leqq 1/(n+1) < 1/n$. Hence,

$$\left| r\frac{p_{n+1}}{P_n} + \frac{r(r-1)}{2!}\frac{p_{n+1}^2}{P_n^2} + \cdots \right|$$

$$\leqq \frac{1}{n}\left(r + \frac{|r(r-1)|}{2}\frac{1}{n} + \cdots + \frac{|r(r-k+1)|}{k!}\frac{1}{n^{k-1}} + \cdots \right) \leqq \frac{A}{n}.$$

Therefore,

$$|P_n^r - P_{n+1}^r| = P_n^r \left| r\frac{p_{n+1}}{P_n} + \frac{r(r-1)}{2!}\frac{p_{n+1}^2}{P_n^2} + \cdots \right| \leqq A\frac{P_n^r}{n},$$

and

$$\left| \frac{P_n^r}{n} - \frac{P_{n+1}^r}{n+1} \right| \leqq \frac{P_n^r}{n(n+1)} + \frac{AP_n^r}{n(n+1)} \leqq A\frac{P_n^r}{n^2}.$$

Consequently,

$$\begin{aligned}
\frac{P_n^r}{n} &= \frac{P_1^r}{1} + \left(\frac{P_2^r}{2} - \frac{P_1^r}{1}\right) + \cdots + \left(\frac{P_n^r}{n} - \frac{P_{n-1}^r}{n-1}\right) \\
&\leqq P_1^r + \left|\frac{P_2^r}{2} - \frac{P_1^r}{1}\right| + \cdots + \left|\frac{P_n^r}{n} - \frac{P_{n-1}^r}{n-1}\right| \\
&\leqq P_1^r + \frac{AP_1^r}{2^2} + \cdots + \frac{AP_{n-1}^r}{(n-1)^2} \\
&= A\sum_{k=1}^{n-1}\frac{P_k^r}{k^2}.
\end{aligned}$$

On taking $r = 2q/(q-2)$, this becomes

$$\frac{P_n^{2q/(q-2)}}{n} \leqq A\sum_{k=1}^{n}\frac{P_k^{2q/(q-2)}}{k} \leqq A.$$

Therefore,

$$\frac{P_n^2}{n^{1-2/q}} = \left(\frac{P_n^{2q/(q-2)}}{n}\right)^{(q-2)/q} \leqq A.$$

(5.45) LEMMA. *If* $\sum_{n=1}^{\infty} P_n^{2q/(q-2)}/n^2 \leqq A$, $q > 2$, *then* $\int_0^{1/n} P^2(t^{-1})\,dt \leqq An^{-2/q}$.

Proof. On making the substitution $t = 1/u$ we get

$$\begin{aligned}
\int_0^{1/n} P^2(t^{-1})\,dt &= \int_n^{\infty} \frac{P^2(u)}{u^2}\,du \\
&\leqq A\sum_{k=n}^{\infty}\frac{P_k^2}{k^2} \\
&= A\sum_{k=n}^{\infty}\frac{P_k^2}{k^{2(q-2)/q}}\frac{1}{k^{4/q}}
\end{aligned}$$

$$\leqq A\left\{\sum_{k=n}^{\infty}\frac{P_k^{2q/(q-2)}}{k^2}\right\}^{(q-2)/q}\left\{\sum_{n}^{\infty}\frac{1}{k^2}\right\}^{2/q} \qquad \text{(Hölder's inequality)}$$

$$\leqq A\left\{\sum_{k=n}^{\infty}\frac{1}{k^2}\right\}^{2/q}$$

$$\leqq An^{-2/q}.$$

Proof of (5.39). From our hypotheses and Lemma (5.42), it is clear that the conditions of Lemmas (5.16) and (5.18) are satisfied. Therefore, the inequalities (5.17) and (5.19) are true. Hence, in view of Lemma (5.43), it suffices to show that

$$\sum_{n=1}^{\infty}\frac{|\alpha_n|+|\beta_n|}{P_{n-1}}\leqq A, \qquad \sum_{n=1}^{\infty}\frac{|\tilde{\alpha}_n|+|\tilde{\beta}_n|}{P_{n-1}}\leqq A.$$

As in Theorem (5.26), we obtain

$$\sum_{n=1}^{\infty}\alpha_n^2\sin^2 nh \leqq A\left\{\int_0^{\pi}\alpha^2(t+h)\,|\,\phi(t+h)-\phi(t-h)\,|^2\,dt + \int_0^{\pi}\phi^2(t-h)\,|\,\alpha(t+h)-\alpha(t-h)\,|^2\,dt\right\}$$

$$= J_1(h)+J_2(h).$$

Now, taking h to be positive, we get

$$J_1(h) = A\int_0^{\pi}\alpha^2(t+h)\,|\,\phi(t+h)-\phi(t-h)\,|^2\,dt$$

$$\leqq A\left\{\int_0^{\pi}|\,\phi(t+h)-\phi(t-h)\,|^q\,dt\right\}^{2/q}\left\{\int_0^{\pi}|\,\alpha(t+h)\,|^{2q/(q-2)}\,dt\right\}^{1-2/q}$$

$$\leqq Ah^{2\alpha}\left\{\int_0^{\pi}|\,\alpha(t+h)\,|^{2q/(q-2)}\,dt\right\}^{1-2/q}$$

$$= Ah^{2\alpha}\left\{\int_h^{\pi+h}|\,\alpha(t)\,|^{2q/(q-2)}\,dt\right\}^{1-2/q}$$

$$\leqq Ah^{2\alpha}\left\{\int_h^{\pi+h}P^{2q/(q-2)}(t^{-1})\right\}^{1-2/q}$$

$$= Ah^{2\alpha}\left\{\int_{(\pi+h)^{-1}}^{h^{-1}}P^{2q/(q-2)}(t)t^{-2}\,dt\right\}^{1-2/q}$$

$$\leqq Ah^{2\alpha}\left\{A+\sum_{k=1}^{[h^{-1}]+1}P_k^{2q/(q-2)}k^{-2}\right\}^{1-2/q}$$

$$\leqq Ah^{2\alpha}.$$

We will write

$$J_2(h) = A\int_0^\pi \phi^2(t-h)\,|\,\alpha(t+h)-\alpha(t-h)\,|^2\,dt$$

$$\leqq A\left\{\int_h^h \phi^2(t)\alpha^2(t+2h)\,dt + \int_{-h}^h \phi^2(t)\alpha^2(t)\,dt + \int_h^\pi \phi^2(t)\,|\,\alpha(t+2h)-\alpha(t)\,|^2\,dt\right\}$$

$$= J_{2,1}(h) + J_{2,2}(h) + J_{2,3}(h).$$

Then we observe first that

$$J_{2,1}(h) \leqq Ah^{2\alpha-2/q}\int_{-h}^h \alpha^2(t+2h)\,dt, \qquad \text{by Lemma (5.41),}$$

$$\leqq Ah^{2\alpha-2/q}\,hP^2(h^{-1})$$

$$\leqq Ah^{2\alpha}, \qquad \text{by Lemma (5.44).}$$

Next,

$$J_{2,2}(h) \leqq Ah^{2\alpha-2/q}\int_0^h P^2(t^{-1})\,dt$$

$$\leqq Ah^{2\alpha-2/q+2/q}, \qquad \text{by Lemma (5.45),}$$

$$= Ah^{2\alpha}.$$

Finally,

$$J_{2,3}(h) \leqq Ah^2P^2(h^{-1})\int_h^\pi t^{-2}\phi^2(t)\,dt, \qquad \text{by Lemma (5.20).}$$

Now,

$$\int_h^\pi t^{-2}\phi^2(t)\,dt \leqq A\int_h^\pi t^{2\alpha-2/q-2}\,dt, \qquad \text{by Lemma (5.41),}$$

$$\leqq \begin{cases} Ah^{2\alpha-2/q-1} & \text{for } 2\alpha-2/q-1<0,\\ A & \text{for } 2\alpha-2/q-1>0,\\ A\log h^{-1} & \text{for } 2\alpha-2/q-1=0.\end{cases}$$

Therefore,

$$J_{2,3}(h) \leqq \begin{cases} Ah^{2\alpha-2/q+1}P^2(h^{-1}) & \text{for } \alpha < 1/q+1/2,\\ Ah^2P^2(h^{-1})\log h^{-1} & \text{for } \alpha = 1/q+1/2,\\ Ah^2P^2(h^{-1}) & \text{for } \alpha > 1/q+1/2,\end{cases}$$

$$\leqq \begin{cases} Ah^{2\alpha} & \text{for } \alpha < 1/q+1/2 \text{ by Lemma (5.44),}\\ Ah^2P^2(h^{-1})\log h^{-1} & \text{for } \alpha = 1/q+1/2,\\ Ah^2P^2(h^{-1}) & \text{for } \alpha > 1/q+1/2.\end{cases}$$

Hence,

$$\sum_{n=1}^{\infty} \alpha_n^2 \sin^2 nh \leqq A(h^{2\alpha} + h^2 P^2(h^{-1}) \log h^{-1}),$$

and the remainder of the theorem follows exactly as in Theorem (5.26).

(5.46) COROLLARY. *If $f(x)$ belongs to Lip (α, q), $1/2 < \alpha \leqq 1$, $q > 2$, then $S(f)$ and $\bar{S}(f)$ are absolutely convergent.*

Proof. Clearly p_n is non-negative and non-increasing, $\lim_{n\to\infty} p_n = 0$, and $|\Delta p_n|$ is non-increasing. Condition (5.39), (1) is satisfied since $P_k = 1$ for all k, and condition (5.39), (2) is satisfied since $\alpha + 1/2 > 1$.

(5.47) COROLLARY. *If $f(x)$ belongs to Lip (α, q), $1/2 \geqq \alpha > 1/q$, $\beta + \alpha > 1/2$, then $S(f)$ and $\bar{S}(f)$ are summable $|C, \beta|$.*

Proof. First we observe that, for given α and q with $\beta + \alpha > 1/2$, we can find $\beta_1 \leqq \beta$ such that $1/2 \geqq \alpha > 1/2 - \beta_1 > 1/q$; hence it suffices to establish our result for such β_1, since, by Theorem (2.20), $|C, \beta_1| \subset |C, \beta|$ for $\beta_1 \leqq \beta$. There will then be no loss of generality in assuming $1/2 \geqq \alpha > 1/2 - \beta > 1/q$. From the last inequality we deduce $\beta < 1/q + 1/2 < 1$. Hence, p_n is non-negative, non-increasing, $\lim_{n\to\infty} p_n = 0$ and $|\Delta p_n|$ is non-increasing. Moreover, $P_n \simeq n^{\beta}$; hence,

$$(1) \quad \sum_{k=1}^{\infty} P_k^{2q/(q-2)} k^{-2} \leqq A \sum_{k=1}^{\infty} k^{(2\beta q/(q-2))-2}.$$

But

$$-\frac{2\beta q}{q-2} + 2 = 2\frac{q(1-\beta)-2}{q-2} > 2\frac{q(1/2+1/q)-2}{q-2} = 1.$$

Therefore,

$$\sum_{k=1}^{\infty} P_k^{2q/(q-2)} k^{-2} \leqq A.$$

Also,

$$(2) \quad \sum_{k=1}^{\infty} P_k^{-1} k^{-\alpha-\frac{1}{2}} \leqq A \sum_{k=1}^{\infty} k^{-\beta-\alpha-\frac{1}{2}}$$

$$\leqq A,$$

since $\alpha + \beta + 1/2 > 1$.

BROWN UNIVERSITY.

BIBLIOGRAPHY

1. S. BERNSTEIN, *Sur la convergence absolue des séries trigonometriques*, Comptes Rendus, Paris, vol. 158(1914), pp. 1661–1663.
2. L. S. BOSANQUET, *The absolute Cesàro summability of a Fourier series*, Proceedings of the London Mathematical Society, (2), vol. 41(1936), pp. 517–528.

3. M. Fekete, *On the absolute summability* (A) *of infinite series*, Proceedings of the Edinburgh Mathematical Society, (2), vol. 3(1932–1933), pp. 132–134.
4. G. H. Hardy and J. E. Littlewood, *A convergence criterion for Fourier series*, Mathematische Zeitschrift, vol. 28(1928), pp. 612–634.
5. J. M. Hyslop, *On the absolute summability of Fourier series*, Proceedings of the London Mathematical Society, (2), vol. 43(1937), pp. 475–483.
6. E. Kogbetliantz, *Sur les séries absolument sommable par la méthode des moyennes arithmétiques*, Bulletin des Sciences Mathématiques, (2), vol. 49(1925), pp. 234–256.
7. F. M. Mears, *Absolute regularity and the Nörlund mean*, Annals of Mathematics, (2), vol. 38(1937), pp. 594–601.
8. N. E. Nörlund, Lunds Universitets Årsskrift, Avd. 2, vol. 16(1919), no. 3.
9. G. Szegö, *Bemerkungen zu einer Arbeit von Herrn Fejér über die Legendreschen Polynome*, Mathematische Zeitschrift, vol. 25(1926), p. 177.
10. J. M. Whittaker, *The absolute summability of Fourier series*, Proceedings of the Edinburgh Mathematical Society, (2), vol. 2(1930–1931), pp. 1–5.
11. G. F. Woronoi, *Extension of the notion of the limit of the sum of terms of an infinite series*, Annals of Mathematics, (2), vol. 33(1932), pp. 422–428.

ASSOCIATED DOUBLE INTEGRAL VARIATION PROBLEMS

By Earl J. Mickle

Introduction

In a paper entitled *Über adjungierte Variationsprobleme und adjungierte Extremalflächen*, Haar [1][1] has given a variation problem associated with a non-parametric double integral variation problem of the type

$$(1) \qquad J[z] = \iint_R F(p, q)\, dx\, dy = \text{min.}, \qquad p = \frac{\partial z}{\partial x}, \qquad q = \frac{\partial z}{\partial y},$$

in such a way that an extremal surface of the problem (1) determines an extremal surface of the associated problem; and a variation problem associated with a parametric double integral variation problem of the type

$$(2) \qquad I[x, y, z] = \iint_G \Phi(A, B, C)\, du\, dv = \text{min.},$$

$$A = \begin{vmatrix} y_u & z_u \\ y_v & z_v \end{vmatrix}, \qquad B = \begin{vmatrix} z_u & x_u \\ z_v & x_v \end{vmatrix}, \qquad C = \begin{vmatrix} x_u & y_u \\ x_v & y_v \end{vmatrix},$$

in such a way that an extremal surface of the problem (2) determines an extremal surface of the associated problem. It is the purpose of this paper to show that the method used by Haar to determine such associated problems can be used to determine a group of such associated problems. With this end in view we give a summary of the results of Haar.

0.1. **Non-parametric adjoint variation problem of Haar.** Let us assume that the integrand function $F(p, q)$ of the problem (1) is of class[2] C'' in a region S of the pq-plane and define the functions

$$(3) \qquad \begin{aligned} X(p, q) &= -F_p(p, q), & Y(p, q) &= -F_q(p, q), \\ Z(p, q) &= F - pF_p - qF_q, & \Delta(p, q) &= F_{pp}F_{qq} - F_{pq}^2. \end{aligned}$$

Let us further assume that the functions $F(p, q)$, $Z(p, q)$, $\Delta(p, q)$ are different from zero everywhere in S and that the transformation[3]

$$T_3: \quad p_3 = -\frac{X(p, q)}{Z(p, q)}, \qquad q_3 = -\frac{Y(p, q)}{Z(p, q)}, \qquad F_3(p_3, q_3) = -\frac{1}{Z(p, q)},$$

Received November 21, 1941.

[1] Numbers in square brackets refer to the bibliography given at the end of this paper.

[2] A function is said to be of class $C^{(n)}$ in a region if the function together with its partial derivatives up to and including those of the n-th order are continuous in the region. A surface $z = z(x,y)$ is said to be of class $C^{(n)}$ in a region of the xy-plane if $z(x,y)$ is of class $C^{(n)}$ in the region.

[3] The particular choice of subscripts here used is for reference in later work.

$$\Delta_3(p, q) = \frac{\partial(p_3, q_3)}{\partial(p, q)} = \frac{F\Delta}{Z^3} \neq 0$$

carries the region S in a one-to-one and continuous way into a region S_3 of the p_3q_3-plane in which the function $F_3(p_3, q_3)$ is defined. Haar called the variation problem

$$(4) \qquad J_3[z_3] = \iint_{R_3} F_3(p_3, q_3)\, dx_3\, dy_3, \qquad p_3 = \frac{\partial z_3}{\partial x_3}, \qquad q_3 = \frac{\partial z_3}{\partial y_3},$$

considered in an (x_3, y_3, z_3)-coordinate system, the adjoint variation problem of (1).

A surface $z = z(x, y)$ of class C'' is called an extremal surface of the problem (1) if $z(x, y)$ satisfies in R the Euler-Lagrange equation[4]

$$(5) \qquad \begin{gathered} F_{pp}(p, q)r + 2F_{pq}(p, q)s + F_{qq}(p, q)t = 0, \\ r = \frac{\partial^2 z}{\partial x^2}, \qquad s = \frac{\partial^2 z}{\partial x \partial y}, \qquad t = \frac{\partial^2 z}{\partial y^2}. \end{gathered}$$

This equation is a necessary and sufficient condition that there exist three single valued auxiliary functions $\xi(x, y)$, $\eta(x, y)$, $\zeta(x, y)$, uniquely determined up to additive constants, which together with the function $z(x, y)$ satisfy in R the system of equations

$$(6) \qquad \begin{aligned} d\xi &= Y(p, q)\, dz - Z(p, q)\, dy, \\ d\eta &= Z(p, q)\, dx - X(p, q)\, dz, \\ d\zeta &= X(p, q)\, dy - Y(p, q)\, dx, \end{aligned}$$

where $dz = p\, dx + q\, dy$. This statement follows from the fact that the cross differentiation test for exact differentials applied to the right sides of (6) reduces to the Euler-Lagrange equation (5).

Haar showed that the transformation

$$x_3 = \xi(x, y), \qquad y_3 = \eta(x, y), \qquad z_3(x_3, y_3) = \zeta(x, y),$$

$t_3[z]$:

$$\delta_3[z] = \frac{\partial(x_3, y_3)}{\partial(x, y)} = FZ \neq 0,$$

$$\xi_3(x_3, y_3) = x, \qquad \eta_3(x_3, y_3) = y, \qquad \zeta_3(x_3, y_3) = z(x, y),$$

where it is assumed that the interior of the region R is carried in a one-to-one and continuous way into the interior of a region R_3 of the x_3y_3-plane, defines four functions $z_3(x_3, y_3)$, $\xi_3(x_3, y_3)$, $\eta_3(x_3, y_3)$, $\zeta_3(x_3, y_3)$ which satisfy the system of equations

$$\begin{aligned} d\xi_3 &= -F_{3q_3}\, dz_3 - (F_3 - p_3F_{3p_3} - q_3F_{3q_3})\, dy_3, \\ d\eta_3 &= (F_3 - p_3F_{3p_3} - q_3F_{3q_3})\, dx_3 + F_{3p_3}\, dz_3, \\ d\zeta_3 &= -F_{3p_3}\, dy_3 + F_{3q_3}\, dx_3. \end{aligned}$$

[4] See, for example, O. Bolza, *Vorlesungen über Variationsrechnung*, pp. 652–656.

That is to say, the surface $z_3 = z_3(x_3, y_3)$ is an extremal surface of the adjoint variation problem (4).

0.2. **Parametric adjoint variation problem of Haar.** Let us assume that the integrand function $\Phi(A, B, C)$ of the variation problem (2) is of class C'' in a region Σ of (A, B, C)-space, that the quantity $(\Phi^4/C^2)(\Phi_{AA}\Phi_{BB} - \Phi_{AB}^2)$ is everywhere different from zero in Σ, and that the transformation

$$\bar{A} = \Phi\Phi_A, \qquad \bar{B} = \Phi\Phi_B, \qquad \bar{C} = \Phi\Phi_C,$$

$$\bar{\Phi}(\bar{A}, \bar{B}, \bar{C}) = -\Phi(A, B, C), \qquad \frac{\partial(\bar{A}, \bar{B}, \bar{C})}{\partial(A, B, C)} = (\Phi^4/C^2)(\Phi_{AA}\Phi_{BB} - \Phi_{AB}^2) \neq 0$$

carries the region Σ in a one-to-one and continuous way into a region $\bar{\Sigma}$ of $(\bar{A}, \bar{B}, \bar{C})$-space in which $\bar{\Phi}(\bar{A}, \bar{B}, \bar{C})$ is defined. Haar called the variation problem

$$\bar{I}[\bar{x}, \bar{y}, \bar{z}] = \iint_G \bar{\Phi}(\bar{A}, \bar{B}, \bar{C})\, du\, dv,$$

(7)
$$\bar{A} = \begin{vmatrix} \bar{y}_u & \bar{z}_u \\ \bar{y}_v & \bar{z}_v \end{vmatrix}, \qquad \bar{B} = \begin{vmatrix} \bar{z}_u & \bar{x}_u \\ \bar{z}_v & \bar{x}_v \end{vmatrix}, \qquad \bar{C} = \begin{vmatrix} \bar{x}_u & \bar{y}_u \\ \bar{x}_v & \bar{y}_v \end{vmatrix},$$

the adjoint variation problem of the problem (2).

A surface

(8)
$$x = x(u, v), \qquad y = y(u, v), \qquad z = z(u, v)$$

of class C'' which satisfies the Euler-Lagrange equations

$$\frac{\partial}{\partial u}(y_v\Phi_C - z_v\Phi_B) + \frac{\partial}{\partial v}(z_u\Phi_B - y_u\Phi_C) = 0,$$

(9)
$$\frac{\partial}{\partial u}(z_v\Phi_A - x_v\Phi_C) + \frac{\partial}{\partial v}(x_u\Phi_C - z_u\Phi_A) = 0,$$

$$\frac{\partial}{\partial u}(x_v\Phi_B - y_v\Phi_A) + \frac{\partial}{\partial v}(y_u\Phi_A - x_u\Phi_B) = 0$$

is called an extremal surface of the variation problem (2).[5] These equations are a necessary and sufficient condition that there exist three single valued auxiliary functions

(10)
$$\bar{x}(u, v), \qquad \bar{y}(u, v), \qquad \bar{z}(u, v)$$

of class C'', uniquely determined up to additive constants, which together with the functions $x(u, v)$, $y(u, v)$, $z(u, v)$ satisfy in G (see equation (2)) the system of equations

[5] See, for example, O. Bolza, loc. cit., pp. 663–667.

$$
\begin{aligned}
d\bar{x} &= \Phi_B\, dz - \Phi_C\, dy, \\
d\bar{y} &= \Phi_C\, dx - \Phi_A\, dz, \\
d\bar{z} &= \Phi_A\, dy - \Phi_B\, dx.
\end{aligned} \tag{11}
$$

Haar called the surface

$$x = \bar{x}(u, v), \qquad y = \bar{y}(u, v), \qquad z = \bar{z}(u, v), \tag{12}$$

determined by the auxiliary functions (10), the adjoint extremal surface of (8) and showed that it is an extremal surface of the adjoint variation problem (7).

0.3. **Extremal surfaces of non-parametric variation problems.** An extremal surface of the problem (1) has been defined as a surface of class C'' which satisfies the Euler-Lagrange equation (5). That is to say, an extremal surface satisfies a condition which must necessarily be satisfied by a surface if it is a minimizing surface of the problem (1). By considering variations on the dependent variable, Haar [2] has shown that if a surface $z = z(x, y)$ of class C' furnishes a minimum for the integral (1) for prescribed boundary conditions then the third equation of (6) must be satisfied by $z(x, y)$ and an auxiliary function $\zeta(x, y)$. By considering variations on each of the independent variables, Radó [3] has shown that if a surface $z = z(x, y)$ of class C' furnishes a minimum for the integral (1) for prescribed boundary conditions, then the first two equations of (6) must be satisfied by $z(x, y)$ and two auxiliary functions $\xi(x, y)$ and $\eta(x, y)$. In this paper we shall call a surface of class C' which satisfies these necessary conditions for a minimizing surface an extremal surface. That is to say, a surface $z = z(x, y)$ of class C' shall be called an extremal surface of the variation problem (1) if there exist three single valued functions $\xi(x, y)$, $\eta(x, y)$, $\zeta(x, y)$, uniquely determined up to additive constants, which together with the function $z(x, y)$ satisfy the system of equations (6). We shall refer to the system of equations (6) as the Haar-Radó system of equations of the variation problem (1) and to the functions ξ, η, ζ as the Haar-Radó auxiliary functions corresponding to a given extremal surface.

0.4. **An associated variation problem related to the Hamiltonian function.** If variables π, κ are introduced by means of the transformation

$$\pi = -X(p, q), \qquad \kappa = -Y(p, q),$$

the Hamiltonian function of the problem (1) is defined by the relation

$$H(\pi, \kappa) = -Z(p, q).$$

By making a transformation similar to the above transformation defining the Hamiltonian function, the author[6] has shown that it is possible to define an

[6] E. J. Mickle, *Hamiltonian and quasi-Hamiltonian functions associated with double integral variation problems*, a doctoral dissertation written under the supervision of Professor Lincoln LaPaz at The Ohio State University in 1941.

associated variation problem in the following way. Assume that $\Delta(p, q)$ is everywhere different from zero in a region S of the pq-plane and that the transformation

$$T_1: \quad p_1 = -Y(p, q), \qquad q_1 = X(p, q), \qquad F_1(p_1, q_1) = Z(p, q),$$
$$\Delta_1(p, q) = \frac{\partial(p_1, q_1)}{\partial(p, q)} = \Delta \neq 0$$

carries the region S in a one-to-one and continuous way into a region S_1 of the p_1q_1-plane in which $F_1(p_1, q_1)$ is defined. If $\xi(x, y)$, $\eta(x, y)$, $\zeta(x, y)$ are the Haar-Radó auxiliary functions corresponding to a given extremal surface $z = z(x, y)$ of the variation problem (1), the transformation

$$t_1[z]: \quad x_1 = x, \qquad y_1 = y, \qquad z_1(x_1, y_1) = \zeta(x, y), \qquad \delta_1[z] = \frac{\partial(x_1, y_1)}{\partial(x, y)} = 1,$$
$$\xi_1(x_1, y_1) = \xi(x, y), \qquad \eta_1(x_1, y_1) = \eta(x, y), \qquad \zeta_1(x_1, y_1) = z(x, y)$$

determines an extremal surface $z_1 = z_1(x_1, y_1)$ of an associated variation problem

$$(13) \qquad J_1[z_1] = \iint_{R_1} F_1(p_1, q_1)\, dx_1\, dy_1, \qquad p_1 = \frac{\partial z_1}{\partial x_1}, \qquad q_1 = \frac{\partial z_1}{\partial y_1},$$

considered in an (x_1, y_1, z_1)-coordinate system, for which $\xi_1(x_1, y_1)$, $\eta_1(x_1, y_1)$, $\zeta_1(x_1, y_1)$ are the corresponding Haar-Radó auxiliary functions.

0.5. **Associated variation problem of Radó.** Subsequent to the completion of the dissertation referred to in §0.4, T. Radó called to the attention of the author that in some unpublished notes he had investigated in addition to the variation problem $J_1[z_1]$ a third associated variation problem defined as follows.[7] Assume that $F(p, q)$ and $Z(p, q)$ are everywhere different from zero in a region S of the pq-plane and that the transformation

$$T_2: \quad p_2 = -\frac{q}{F(p, q)}, \qquad q_2 = \frac{p}{F(p, q)}, \qquad F_2(p_2, q_2) = -\frac{1}{F(p, q)},$$
$$\Delta_2(p, q) = \frac{\partial(p_2, q_2)}{\partial(p, q)} = \frac{Z}{F^2} \neq 0$$

carries the region S in a one-to-one and continuous way into a region S_2 of the p_2q_2-plane in which $F_2(p_2, q_2)$ is defined. If $\xi(x, y)$, $\eta(x, y)$, $\zeta(x, y)$ are the Haar-Radó auxiliary functions corresponding to a given extremal surface $z = z(x, y)$ of the problem (1), the transformation

$$t_2[z]: \quad x_2 = \xi(x, y), \qquad y_2 = \eta(x, y), \qquad z_2(x_2, y_2) = z(x, y),$$
$$\delta_2[z] = \frac{\partial(x_2, y_2)}{\partial(x, y)} = FZ \neq 0,$$
$$\xi_2(x_2, y_2) = x, \qquad \eta_2(x_2, y_2) = y, \qquad \zeta_2(x_2, y_2) = \zeta(x, y),$$

[7] The author wishes to express his appreciation to Professor T. Radó for the use of these notes and for his suggestions during the preparation of this paper.

where it is assumed that the interior of the region R is carried in a one-to-one and continuous way into the interior of a region R_2 of the x_2y_2-plane, determines an extremal surface $z_2 = z_2(x_2, y_2)$ of an associated variation problem

$$J_2[z_2] = \iint_{R_2} F_2(p_2, q_2)\, dx_2\, dy_2, \qquad p_2 = \frac{\partial z_2}{\partial x_2}, \qquad q_2 = \frac{\partial z_2}{\partial y_2}, \tag{14}$$

considered in an (x_2, y_2, z_2)-coordinate system, for which $\xi_2(x_2, y_2)$, $\eta_2(x_2, y_2)$, $\zeta_2(x_2, y_2)$ are the corresponding Haar-Radó auxiliary functions.

0.6. **Some further associated variation problems.** The transformations T_i, $i = 1, 2, 3$, defining the variables p_i, q_i and the functions $F_i(p_i, q_i)$ are of the form

$$\begin{gathered} p_i = \alpha_i(X, Y, Z, p, q, F), \qquad q_i = \beta_i(X, Y, Z, p, q, F), \\ F_i(p_i, q_i) = \gamma_i(X, Y, Z, p, q, F), \qquad \Delta_i(p, q) = \frac{\partial(p_i, q_i)}{\partial(p, q)} \neq 0 \end{gathered} \tag{15}$$

and the transformations $t_i[z]$, $i = 1, 2, 3$, are of the form

$$\begin{gathered} x_i = a_i(x, y, z, p, q), \quad y_i = b_i(x, y, z, p, q), \quad z_i(x_i, y_i) = c_i(x, y, z, p, q), \\ \xi_i(x_i, y_i) = d_i(x, y, z, p, q), \quad \eta_i(x_i, y_i) = e_i(x, y, z, p, q), \\ \zeta_i(x_i, y_i) = f_i(x, y, z, p, q), \\ \delta_i[z] = \partial(x_i, y_i)/\partial(x, y) \neq 0, \end{gathered} \tag{16}$$

where a_i, b_i, c_i, d_i, e_i, f_i are in some order the six functions

$$\begin{gathered} x, y, z, \qquad \int_{x^1,y^1}^{x,y} Y\, dz - Z\, dy, \qquad \int_{x^1,y^1}^{x,y} Z\, dx - X\, dz, \\ \int_{x^1,y^1}^{x,y} X\, dy - Y\, dx \end{gathered} \tag{17}$$

with $dz = p\, dx + q\, dy$. When evaluated on an extremal surface $z = z(x, y)$ of the variation problem (1), the functions (17) reduce to the functions

$$x, y, z(x, y), \xi(x, y), \eta(x, y), \zeta(x, y), \tag{18}$$

where ξ, η, ζ are the corresponding Haar-Radó auxiliary functions. The question arises as to whether there are any further transformations of the type (15) defining an associated variation problem and a corresponding transformation of the type (16), using the six functions (17), such that the transformation $t_i[z]$ when evaluated on an extremal surface of the problem (1) determines an extremal surface of the associated problem and the corresponding Haar-Radó auxiliary functions. In Part I we give twenty-four such transformations, including the three already mentioned, which fulfill these conditions. The transformations T_i have the further property that they form a group of order twenty-four.

In Part II we give eight variation problems, including the adjoint variation problem of Haar, such that each of the eight surfaces

$$(19) \qquad x = \begin{cases} x(u, v) \\ \text{or} \\ \bar{x}(u, v), \end{cases} \qquad y = \begin{cases} y(u, v) \\ \text{or} \\ \bar{y}(u, v), \end{cases} \qquad z = \begin{cases} z(u, v) \\ \text{or} \\ \bar{z}(u, v), \end{cases}$$

where $x(u, v)$, $y(u, v)$, $z(u, v)$, $\bar{x}(u, v)$, $\bar{y}(u, v)$, $\bar{z}(u, v)$ are functions satisfying the system of equations (11), is an extremal surface of one of these associated problems. The eight transformations defining the associated variation problems also form a group.

PART I

Associated Non-parametric Variation Problems

1.1. **Twenty-four associated variation problems.** We shall assume that the integrand function $F(p, q)$ of the variation problem (1) is of class $C^{(n)}$, $n \geqq 2$, in an open region S of the pq-plane. Consider the twenty-four transformations

$$(1.1) \qquad T_i: \quad p_i = -Y_j(p, q), \quad q_i = X_j(p, q), \quad F_i(p_i, q_i) = Z_j(p, q),$$

where $j = i + (-1)^i$, $i = 0, 1, 2, \cdots, 23$, and where the functions $X_j(p, q)$, $Y_j(p, q)$, $Z_j(p, q)$ are defined in Table I below in terms of the functions $X(p, q)$, $Y(p, q)$, $Z(p, q)$ given in (3) and the functions p, q and $F(p, q)$. We shall assume for each particular transformation that in the region S the quantity

$$(1.2) \qquad \Delta_i(p, q) = \frac{\partial(p_i, q_i)}{\partial(p, q)} = X_{jp}Y_{jq} - X_{jq}Y_{jp}, \qquad j = i + (-1)^i,$$

is everywhere different from zero and that the transformation T_i carries the region S in a one-to-one and continuous way into a region S_i of the p_iq_i-plane in which $F_i(p_i, q_i)$ is defined. The values of the $\Delta_i(p, q)$ are given in Table I where $\Delta(p, q)$ has the value given in (3).

TABLE I

i	0	2	4	6	8	10	12	14	16	18	20	22
X_i	X	$-\frac{Y}{Z}$	Z	$-\frac{1}{X}$	$\frac{Y}{X}$	Y	$\frac{Z}{Y}$	$-\frac{1}{Y}$	$-\frac{1}{Z}$	$\frac{X}{Z}$	$-\frac{X}{Y}$	$-\frac{Z}{X}$
Y_i	Y	$\frac{X}{Z}$	X	$-\frac{Z}{X}$	$-\frac{1}{X}$	Z	$-\frac{1}{Y}$	$-\frac{X}{Y}$	$-\frac{Y}{Z}$	$-\frac{1}{Z}$	$\frac{Z}{Y}$	$\frac{Y}{X}$
Z_i	Z	$-\frac{1}{Z}$	Y	$\frac{Y}{X}$	$-\frac{Z}{X}$	X	$-\frac{X}{Y}$	$\frac{Z}{Y}$	$\frac{X}{Z}$	$-\frac{Y}{Z}$	$-\frac{1}{Y}$	$-\frac{1}{X}$
Δ_i	1	$\frac{Z}{F^3}$	$-\frac{1}{q^3}$	$-\frac{X}{q^3}$	$-X$	$-\frac{1}{p^3}$	$-\frac{Y}{p^3}$	Y	$-\frac{Z}{p^3}$	$\frac{Z}{q^3}$	$\frac{Y}{F^3}$	$\frac{X}{F^3}$

TABLE I (*continued*)

i	1	3	5	7	9	11	13	15	17	19	21	23
X_i	q	$\frac{p}{F}$	$-\frac{p}{q}$	$-\frac{1}{q}$	F	$\frac{1}{p}$	$-\frac{F}{p}$	$-p$	$\frac{q}{p}$	$-\frac{F}{q}$	$-\frac{1}{F}$	$\frac{q}{F}$
Y_i	$-p$	$\frac{q}{F}$	$-\frac{1}{q}$	$-\frac{F}{q}$	q	$\frac{q}{p}$	$\frac{1}{p}$	F	$-\frac{F}{p}$	$-\frac{p}{q}$	$\frac{p}{F}$	$-\frac{1}{F}$
Z_i	F	$-\frac{1}{F}$	$-\frac{F}{q}$	$-\frac{p}{q}$	$-p$	$-\frac{F}{p}$	$\frac{q}{p}$	q	$\frac{1}{p}$	$-\frac{1}{q}$	$\frac{q}{F}$	$\frac{p}{F}$
Δ_i	Δ	$\frac{F\Delta}{Z^2}$	$-q\Delta$	$-\frac{q\Delta}{X^2}$	$-\frac{\Delta}{X^2}$	$-p\Delta$	$-\frac{p\Delta}{Y^2}$	$\frac{\Delta}{Y^2}$	$-\frac{p\Delta}{Z^2}$	$\frac{q\Delta}{Z^2}$	$\frac{F\Delta}{Y^2}$	$\frac{F\Delta}{X^2}$

The equations defining F_{ip_i}, F_{iq_i}, $F_i - p_iF_{ip_i} - q_iF_{iq_i}$ can be found as follows. Since in the transformation T_i, $Y_j(p, q)$ and $X_j(p, q)$ are of class $C^{(n-1)}$, it follows from implicit function existence theorems that p and q as functions p_i and q_i are of class $C^{(n-1)}$. Therefore,

$$dp_i = -Y_{ip}\,dp - Y_{iq}\,dq; \qquad dq_i = X_{ip}\,dp + X_{iq}\,dq.$$

From these equations,

$$dp = \frac{X_{jq}}{\Delta_i}\,dp_i + \frac{Y_{jq}}{\Delta_i}\,dq_i; \qquad dq = -\frac{X_{jp}}{\Delta_i}\,dp_i - \frac{Y_{jp}}{\Delta_i}\,dq_i.$$

Thus,

$$\frac{\partial p}{\partial p_i} = \frac{X_{jq}}{\Delta_i}; \qquad \frac{\partial p}{\partial q_i} = \frac{Y_{jq}}{\Delta_i}; \qquad \frac{\partial q}{\partial p_i} = -\frac{X_{jp}}{\Delta_i}; \qquad \frac{\partial q}{\partial q_i} = -\frac{Y_{jp}}{\Delta_i}$$

and hence

$$F_{ip_i} = Z_{jp}\frac{\partial p}{\partial p_i} + Z_{jq}\frac{\partial q}{\partial p_i} = (X_{jq}Z_{jp} - X_{jp}Z_{jq})/\Delta_i;$$

(1.3)

$$F_{iq_i} = Z_{ip}\frac{\partial p}{\partial q_i} + Z_{jq}\frac{\partial q}{\partial q_i} = (Y_{jq}Z_{jp} - Y_{jp}Z_{jq})/\Delta_i.$$

From Table I and the relation (1.2), we obtain

(1.4) $\qquad F_{ip_i} = -X_i, \quad F_{iq_i} = -Y_i, \quad F_i - p_iF_{ip_i} - q_iF_{iq_i} = Z_i.$

For example, for $i = 3$, $X_2 = Y_{16}$, $Z_2 = X_{16}$, $Y_2 = X_{18}$, $Z_2 = Y_{18}$. Substituting these values in (1.3) and using (1.2) and Table I gives

$$F_{3p_3} = \frac{\Delta_{17}}{\Delta_3} = -\frac{p}{F} = -X_3, \qquad F_{3q_3} = \frac{\Delta_{19}}{\Delta_3} = -\frac{q}{F} = -Y_3.$$

From these equations and the transformation T_3,

$$F_3 - p_3F_{3p_3} - q_3F_{3q_3} = Z_2 - Y_2X_3 + X_2Y_3$$
$$= -(F + pX + qY)/FZ = -1/F = Z_3.$$

The equations (1.4) show that $F_i(p_i, q_i)$ is of class $C^{(n)}$ in S_i.

We shall call the variation problem

$$(1.5) \qquad J_i[z_i] = \iint_{R_i} F_i(p_i, q_i)\, dx_i\, dy_i, \qquad p_i = \frac{\partial z_i}{\partial x_i}, \qquad q_i = \frac{\partial z_i}{\partial y_i},$$

considered in an (x_i, y_i, z_i)-coordinate system, the i-th associated variation problem of the original variation problem (1).

We note that the problem $J_0[z_0]$ is the problem (1) itself, the problem $J_1[z_1]$ is the problem given in §0.4, the problem $J_2[z_2]$ is the associated variation problem of Radó, and the problem $J_3[z_3]$ is the adjoint variation problem of Haar.

1.2. **Extremal surfaces of the associated problems.** In Table II below we give twenty-four transformations of the type (16) evaluated on an extremal surface $z = z(x, y)$ of the problem (1) with $\xi(x, y)$, $\eta(x, y)$, $\zeta(x, y)$ as the corresponding Haar-Radó auxiliary functions. The assumption that $\Delta_i(p, q) \neq 0$ in S, that the transformation T_i determines a one-to-one and continuous mating between S and S_i, and that the function $F_i(p_i, q_i)$ is defined for every point of S_i implies that $\delta_i[z] = \dfrac{\partial(x_i, y_i)}{\partial(x, y)} \neq 0$ in R for $\left(\dfrac{\partial z}{\partial x}, \dfrac{\partial z}{\partial y}\right)$ in S. In Table II it is to be noted that $\delta_i[z] = \delta_j[z]$ for $j = i + (-1)^i$. We shall assume in each case that the transformation $t_i[z]$ carries the interior of the region R of the xy-plane in a one-to-one and continuous way into the interior of a region R_i of the x_iy_i-plane in which the functions $z_i(x_i, y_i)$, $\xi_i(x_i, y_i)$, $\eta_i(x_i, y_i)$, $\zeta_i(x_i, y_i)$ are defined. From implicit function existence theorems, x and y as functions of x_i and y_i are of class C' and hence so are z_i, ξ_i, η_i and ζ_i.

TABLE II

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
$x_i =$	x	x	ξ	ξ	z	z	x	x	ζ	ζ	y	y	ξ	ξ	y	y	z	z	η	η	ζ	ζ	η	η
$y_i =$	y	y	η	η	x	x	η	η	x	x	z	z	y	y	ζ	ζ	ξ	ξ	z	z	ξ	ξ	ζ	ζ
$z_i(x_i, y_i) =$	z	ζ	z	ζ	y	η	ζ	z	η	y	x	ξ	ζ	z	ξ	x	η	y	ξ	x	y	η	x	ξ
$\xi_i(x_i, y_i) =$	ξ	ξ	x	x	ζ	ζ	ξ	ξ	z	z	η	η	x	x	η	η	ζ	ζ	y	y	z	z	y	y
$\eta_i(x_i, y_i) =$	η	η	y	y	ξ	ξ	y	y	ξ	ξ	ζ	ζ	η	η	z	z	x	x	ζ	ζ	x	x	z	z
$\zeta_i(x_i, y_i) =$	ζ	z	ζ	z	η	y	z	ζ	y	η	ξ	x	z	ζ	x	ξ	y	η	x	ξ	η	y	ξ	x
$\delta_i[z] =$	1		FZ		$-q$		$-qX$		$-X$		$-p$		pY		Y		$-pZ$		qZ		FY		FX	

The Haar-Radó system of equations (6) can be written

$$(1.6) \qquad \begin{aligned} dz &= p\,dx + q\,dy, \\ d\xi &= Y\,dz - Z\,dy = -p\,d\zeta - F\,dy, \\ d\eta &= Z\,dx - X\,dz = F\,dx - q\,d\zeta, \\ d\zeta &= X\,dy - Y\,dx. \end{aligned}$$

If the values of dx, dy, dz, $d\xi$, $d\eta$, $d\zeta$ as determined by the transformation $t_i[z]$ are substituted in (1.6), this system of equations can be rewritten in the form

$$(i)\qquad \begin{aligned} dz_i &= X_j\,dy_i - Y_j\,dx_i, \qquad j = i + (-1)^i,\\ d\xi_i &= Y_i\,dz_i - Z_i\,dy_i,\\ d\eta_i &= Z_i\,dx_i - X_i\,dz_i,\\ d\zeta_i &= X_i\,dy_i - Y_i\,dx_i, \end{aligned}$$

where, as defined in Table I, X_i, Y_i, Z_i, X_j and Y_j are functions of p and q and are evaluated on the extremal surface $z = z(x, y)$ defining the transformation $t_i[z]$ with x and y as functions of x_i and y_i.

For example, if $i = 3$, then from Table II, $dx = d\xi_3$, $dy = d\eta_3$, $dz = d\zeta_3$, $d\xi = dx_3$, $d\eta = dy_3$, $d\zeta = dz_3$. Substituting these relations in (1.6) gives

$$(1.7)\qquad \begin{aligned} d\zeta_3 &= p\,d\xi_3 + q\,d\eta_3,\\ dx_3 &= -p\,dz_3 - F\,d\eta_3,\\ dy_3 &= F\,d\xi_3 - q\,dz_3,\\ dz_3 &= X\,d\eta_3 - Y\,d\xi_3. \end{aligned}$$

From the third and second equations of (1.7) we obtain respectively on using Table I

$$d\xi_3 = \frac{q}{F}\,dz_3 + \frac{1}{F}\,dy_3 = Y_3\,dz_3 - Z_3\,dy_3,$$

$$d\eta_3 = -\frac{1}{F}\,dx_3 - \frac{p}{F}\,dz_3 = Z_3\,dx_3 - X_3\,dz_3.$$

Substituting these values of $d\xi_3$ and $d\eta_3$ in the first and the fourth equations of (1.7) gives respectively on using Table I

$$\begin{aligned} d\zeta_3 &= p\left(\frac{q}{F}\,dz_3 + \frac{1}{F}\,dy_3\right) - q\left(\frac{1}{F}\,dx_3 + \frac{p}{F}\,dz_3\right) = \frac{p}{F}\,dy_3 - \frac{q}{F}\,dx_3\\ &= X_3\,dy_3 - Y_3\,dx_3,\\ dz_3 &= -X\left(\frac{1}{F}\,dx_3 + \frac{p}{F}\,dz_3\right) - Y\left(\frac{q}{F}\,dz_3 + \frac{1}{F}\,dy_3\right). \end{aligned}$$

From this last equation we obtain

$$(F + pX + qY)\,dz_3 = -Y\,dy_3 - X\,dx_3,$$

which gives, since $Z = F + pX + qY$,

$$dz_3 = -\frac{Y}{Z}\,dy_3 - \frac{X}{Z}\,dx_3 = X_2\,dy_3 - Y_2\,dx_3.$$

If when p and q are functions of x_i and y_i the system of equations (i) is satisfied by four functions $z_i(x_i, y_i)$, $\xi_i(x_i, y_i)$, $\eta_i(x_i, y_i)$, $\zeta_i(x_i, y_i)$, the inverse of the

transformation $t_i[z]$ will determine four functions $z(x, y)$, $\xi(x, y)$, $\eta(x, y)$, $\zeta(x, y)$ which satisfy the system of equations (1.6). That is to say, the surface $z = z(x, y)$ is an extremal surface of the variation problem (1) on which $\delta_i[z] \neq 0$ and for which ξ, η, ζ are the corresponding Haar-Radó auxiliary functions.

If the relations given by the transformation T_i and (1.4) are substituted in the system of equations (i) this system may be rewritten in the form

$$(i') \qquad \begin{aligned} dz_i &= p_i\,dx_i + q_i\,dy_i, \\ d\xi_i &= -F_{iq_i}\,dz_i - (F_i - p_iF_{ip_i} - q_iF_{iq_i})\,dy_i, \\ d\eta_i &= (F_i - p_iF_{ip_i} - q_iF_{iq_i})\,dx_i + F_{ip_i}\,dz_i, \\ d\zeta_i &= -F_{ip_i}\,dy_i + F_{iq_i}\,dx_i. \end{aligned}$$

This system of equations is the Haar-Radó system of equations for the i-th associated problem. These results may be summarized as follows.

The transformation $t_i[z]$ taken on an extremal surface $z = z(x, y)$ of the variation problem (1) *on which $\delta_i[z] \neq 0$ determines an extremal surface $z_i = z_i(x_i, y_i)$ of the i-th associated problem $J_i[z_i]$ for which $\xi_i(x_i, y_i)$, $\eta_i(x_i, y_i)$, $\zeta_i(x_i, y_i)$ are the corresponding Haar-Radó auxiliary functions and, conversely, the inverse of the transformation $t_i[z]$ taken on an extremal surface $z_i = z_i(x_i, y_i)$ of the i-th associated variation problem determines an extremal surface $z = z(x, y)$ of the original problem* (1) *on which $\delta_i[z] \neq 0$ and for which $\xi(x, y)$, $\eta(x, y)$, $\zeta(x, y)$ are the corresponding Haar-Radó auxiliary functions.*

1.3. **Group property of the transformations T_i.** In §1.1 each transformation T_i is considered independently of the other twenty-three transformations. In this section we shall assume that in the region S the quantities $\Delta_i(p, q)$ are everywhere different from zero simultaneously. From Table I this is seen to be equivalent to the assumption that in S the quantities X, Y, Z, F, p, q, Δ are everywhere different from zero. From Table I, the transformations T_i and the relations (1.4), this assumption is seen to imply that in the regions S_i

$$(1.8) \qquad \begin{aligned} &X_i \neq 0, \quad Y_i \neq 0, \quad Z_i \neq 0, \quad p_i \neq 0, \quad q_i \neq 0, \quad F_i \neq 0, \\ &F_{ip_ip_i}F_{iq_iq_i} - F^2_{ip_iq_i} = \Delta_j \neq 0, \quad j = i + (-1)^i,\ i = 0, 1, 2, \cdots, 23. \end{aligned}$$

The functions $X_i(p, q)$, $Y_i(p, q)$, $Z_i(p, q)$ defined in Table I are of the form

$$X_i = \alpha_i(X, Y, Z, p, q, F), \quad Y_i = \beta_i(X, Y, Z, p, q, F), \quad Z_i = \gamma_i(X, Y, Z, p, q, F).$$

By the transformation T_kT_i, we shall mean the transformation

$$(1.9) \qquad \begin{aligned} p_{k,i} &= -\beta_j(X_k, Y_k, Z_k, p_k, q_k, F_k), \\ q_{k,i} &= \alpha_j(X_k, Y_k, Z_k, p_k, q_k, F_k), \qquad j = i + (-1)^i, \\ F_{k,i} &= \gamma_j(X_k, Y_k, Z_k, p_k, q_k, F_k). \end{aligned}$$

By (1.8) this transformation is always possible. We leave for the reader the verification of the following relations:

(1.10) $T_0T_i = T_iT_0 = T_i, \qquad i = 0, 1, 2, \cdots, 23,$

(1.11) $T_i^2 = T_0, \qquad i = 1, 2, 3, 13,$

(1.12) $T_1T_2 = T_3, \qquad T_1T_3 = T_2, \qquad T_2T_3 = T_1,$

(1.13) $T_{13}T_iT_{13} = T_i, \qquad i = 0, 1, 2, 3,$

(1.14) $T_{13}T_i = T_k, \qquad k = 13, 12, 7, 6$ for $i = 0, 1, 2, 3$ respectively,

(1.15) $T_4^2 = T_{10}, T_4^3 = T_0,$

(1.16) $T_{10}T_iT_4 = T_k, \qquad k = 13, 6, 2, 7$ for $i = 1, 2, 3, 13$ respectively,

(1.17) $T_4T_i = T_k, \qquad k = 5, 20, 21, 16, 17, 8, 9$ for $i = 1, 2, 3, 6, 7, 12, 13$ respectively,

(1.18) $T_{10}T_i = T_k, \qquad k = 11, 22, 23, 14, 15, 18, 19$ for $i = 1, 2, 3, 6, 7, 12, 13$ respectively.

From (1.10), (1.11), (1.12) it follows that the transformations T_i, $i = 0, 1, 2, 3$, form a group of order four which we shall call G_1. From (1.11) $T_{13}^{-1} = T_{13}$ and from (1.13) the group G_1 is invariant under the transformation T_{13}. Therefore, G_1 and T_{13} generate a group G_2 of order eight whose elements by (1.14) are T_i, $i = 0, 1, 2, 3, 6, 7, 12, 13$. From (1.15) the inverse of the transformation T_4 is T_{10}. By (1.16) the group G_2 is invariant under the transformation T_4. Therefore, G_2 and T_4 generate a group of order twenty-four whose elements by (1.17) and (1.18) are the transformations T_i, $i = 0, 1, 2, \cdots, 23$.

1.4. **Variation problems associated with the Dirichlet and the area integrals.** The variation problems associated with the Dirichlet integral

(1.19) $$J[z] = \frac{1}{2}\iint_R (p^2 + q^2)\,dx\,dy$$

by means of the transformations T_i are given below. The subscripts have been dropped from the x_i, y_i, z_i, p_i, q_i. The numbers after the problems refer to the transformations T_i giving the associated problem.

(a) $$J[z] = \frac{1}{2}\iint_R (p^2 + q^2)\,dx\,dy, \qquad i = 0, 1, 2, 3;$$

(b) $$J[z] = \iint_R \frac{1 + q^2}{2p}\,dx\,dy, \qquad p \neq 0, i = 4, 9, 16, 21;$$

(c) $$J[z] = \iint_R \frac{1 + p^2}{2q}\,dx\,dy, \qquad q \neq 0, i = 10, 15, 18, 23;$$

(d) $$J[z] = \iint_R (2q - p^2)^{\frac{1}{2}}\,dx\,dy, \qquad 2q - p^2 > 0,\ i = 8, 17;$$

(e) $$J[z] = \iint_R (2p - q^2)^{\frac{1}{2}}\,dx\,dy, \qquad 2p - q^2 > 0,\ i = 11, 22;$$

(f) $$J[z] = \iint_R (2pq - 1)^{\frac{1}{2}}\,dx\,dy, \qquad 2pq > 1,\ i = 7, 12, 13;$$

(g) $$J[z] = \iint_R (-2q - p^2)^{\frac{1}{2}}\,dx\,dy, \qquad 2q + p^2 < 0,\ i = 5, 20;$$

(h) $$J[z] = \iint_R (-2p - q^2)^{\frac{1}{2}}\,dx\,dy, \qquad 2p + q^2 < 0,\ i = 14, 19;$$

(i) $$J[z] = \iint_R (-2pq - 1)^{\frac{1}{2}}\,dx\,dy, \qquad 2pq < -1,\ i = 6.$$

Similarly the variation problems associated with the area integral

(1.20) $$J[z] = \iint_R (1 + p^2 + q^2)^{\frac{1}{2}}\,dx\,dy$$

are

(j) $$J[z] = \iint_R (1 + p^2 + q^2)^{\frac{1}{2}}\,dx\,dy, \qquad i = 0, 3, 4, 10, 21, 23;$$

(k) $$J[z] = \iint_R (1 - p^2 - q^2)^{\frac{1}{2}}\,dx\,dy, \qquad p^2 + q^2 < 1,\ i = 1, 2, 5, 11, 20, 22;$$

(l) $$J[z] = \iint_R (p^2 - q^2 - 1)^{\frac{1}{2}}\,dx\,dy, \qquad p^2 - q^2 > 1,\ i = 6, 9, 13, 14, 16, 19;$$

(m) $$J[z] = \iint_R (q^2 - p^2 - 1)^{\frac{1}{2}}\,dx\,dy, \qquad q^2 - p^2 > 1,\ i = 7, 8, 12, 15, 17, 18.$$

If the integrand function $F(p, q)$ of the problem (1) is an analytic function of p and q in the region S and if an extremal surface $z = z(x, y)$ defining the transformation $t_i[z]$ is an analytic function of x and y in the interior of the region R, then it follows from implicit function existence theorems that the extremal surface $z_i = z_i(x_i, y_i)$ of the i-th associated problem is an analytic function of x_i and y_i in the interior of the region R_i. Haar [2] has shown that an extremal surface of the Dirichlet integral (1.19) is necessarily analytic and Radó [3] has

shown that an extremal surface of the area integral (1.20) is necessarily analytic. We thus have the following results.

If a surface $z = z(x, y)$ of class C' is an extremal surface of one of the variation problems associated with the Dirichlet integral or the area integral, then $z(x, y)$ is an analytic function of x and y in the interior of the region R and in R satisfies the corresponding Euler-Lagrange second order partial differential equation

(a) $r + t = 0$;

(b) $(1 + q^2)r - 2pqs + p^2t = 0, \quad p \neq 0$;

(c) $q^2r - 2pqs + (1 + p^2)t = 0, \quad q \neq 0$;

(d) $2qr - 2ps + t = 0, \quad 2q - p^2 > 0$;

(e) $r - 2qs + 2pt = 0, \quad 2p - q^2 > 0$;

(f) $q^2r - 2(pq - 1)s + p^2t = 0, \quad 2pq > 1$;

(g) $2qr - 2ps - t = 0, \quad 2q + p^2 < 0$;

(h) $r + 2qs - 2pt = 0, \quad 2p + q^2 < 0$;

(i) $q^2r - 2(pq + 1)s + p^2t = 0, \quad 2pq < -1$;

(j) $(1 + q^2)r - 2pqs + (1 + p^2)t = 0$;

(k) $(q^2 - 1)r - 2pqs + (p^2 - 1)t = 0, \quad p^2 + q^2 < 1$;

(l) $(q^2 + 1)r - 2pqs + (p^2 - 1)t = 0, \quad p^2 - q^2 > 1$;

(m) $(q^2 - 1)r - 2pqs + (p^2 + 1)t = 0, \quad q^2 - p^2 > 1$.

PART II

Associated Parametric Variation Problems

2.1. **Some fundamental relations.** For the variation problem

$$I[x, y, z] = \iint_{\sigma} \Phi(A, B, C)\, du\, dv = \text{min.}, \tag{2.1}$$

where

$$A = \begin{vmatrix} y_u & z_u \\ y_v & z_v \end{vmatrix}, \quad B = \begin{vmatrix} z_u & x_u \\ z_v & x_v \end{vmatrix}, \quad C = \begin{vmatrix} x_u & y_u \\ x_v & y_v \end{vmatrix}, \tag{2.2}$$

we make the assumptions

(a) Φ is of class $C^{(n)}$, $n \geqq 2$, in a star shaped region Σ of (A, B, C)-space.[8]

[8] A region Σ of (A, B, C)-space is said to be star shaped if for every number $k > 0$ the point (kA, kB, kC) is in Σ if the point (A, B, C) is in Σ.

(b) Φ is homogeneous of degree one with respect to its three arguments, i.e., for all $k > 0$,

$$\Phi(kA, kB, kC) = k\Phi(A, B, C).$$

The condition (b) implies that

$$\Phi = A\Phi_A + B\Phi_B + C\Phi_C . \tag{2.3}$$

Let

$$x = x(u, v), \qquad y = y(u, v), \qquad z = z(u, v) \tag{2.4}$$

be an arbitrary extremal surface of the variation problem (2.1) and let

$$\bar{x}(u, v), \qquad \bar{y}(u, v), \qquad \bar{z}(u, v) \tag{2.5}$$

be three auxiliary functions of class C'' which together with the functions $x(u, v)$, $y(u, v)$, $z(u, v)$ satisfy the system of equations

$$\begin{aligned} d\bar{x} &= \Phi_B\, dz - \Phi_C\, dy, \\ d\bar{y} &= \Phi_C\, dx - \Phi_A\, dz, \\ d\bar{z} &= \Phi_A\, dy - \Phi_B\, dx. \end{aligned} \tag{2.6}$$

Since A, B, C are direction components of the normals to the surface (2.4), we have

$$A\, dx + B\, dy + C\, dz = 0. \tag{2.7}$$

If in (2.6) the first equation is multiplied by Φ_A, the second by Φ_B, the third by Φ_C, we obtain on adding

$$\Phi_A\, d\bar{x} + \Phi_B\, d\bar{y} + \Phi_C\, d\bar{z} = 0. \tag{2.8}$$

If in (2.6) the first equation is multiplied by B and the second by $-A$, we obtain on adding and using relations (2.3), (2.7)

$$B\, d\bar{x} - A\, d\bar{y} - \Phi\, dz = 0. \tag{2.9}$$

Similarly,

$$C\, d\bar{x} + \Phi\, dy - A\, d\bar{z} = 0; \tag{2.10}$$

$$\Phi\, dx - C\, d\bar{y} + B\, d\bar{z} = 0. \tag{2.11}$$

From (2.6) we have

$$\begin{aligned} \bar{x}_u &= \Phi_B z_u - \Phi_C y_u, \qquad \bar{y}_u = \Phi_C x_u - \Phi_A z_u, \qquad \bar{z}_u = \Phi_A y_u - \Phi_B x_u; \\ \bar{x}_v &= \Phi_B z_v - \Phi_C y_v, \qquad \bar{y}_v = \Phi_C x_v - \Phi_A z_v, \qquad \bar{z}_v = \Phi_A y_v - \Phi_B x_v. \end{aligned} \tag{2.12}$$

From (2.12) we can compute the further relations

$$
\begin{aligned}
&\begin{vmatrix} z_u & \bar{x}_u \\ z_v & \bar{x}_v \end{vmatrix} = A\Phi_C, \quad \begin{vmatrix} \bar{x}_u & y_u \\ \bar{x}_v & y_v \end{vmatrix} = -A\Phi_B, \quad \begin{vmatrix} \bar{y}_u & z_u \\ \bar{y}_v & z_v \end{vmatrix} = -B\Phi_C, \\
&\begin{vmatrix} x_u & \bar{y}_u \\ x_v & \bar{y}_v \end{vmatrix} = B\Phi_A, \quad \begin{vmatrix} y_u & \bar{z}_u \\ y_v & \bar{z}_v \end{vmatrix} = C\Phi_B, \quad \begin{vmatrix} \bar{z}_u & x_u \\ \bar{z}_v & x_v \end{vmatrix} = -C\Phi_A, \\
&\begin{vmatrix} \bar{y}_u & \bar{z}_u \\ \bar{y}_v & \bar{z}_v \end{vmatrix} = \Phi\Phi_A, \quad \begin{vmatrix} \bar{z}_u & \bar{x}_u \\ \bar{z}_v & \bar{x}_v \end{vmatrix} = \Phi\Phi_B, \quad \begin{vmatrix} \bar{x}_u & \bar{y}_u \\ \bar{x}_v & \bar{y}_v \end{vmatrix} = \Phi\Phi_C.
\end{aligned} \tag{2.13}
$$

2.2. **Transformations using functions homogeneous of degree one.** If $\alpha(A, B, C)$, $\beta(A, B, C)$, $\gamma(A, B, C)$, $\delta(A, B, C)$ are four functions homogeneous of degree one and if the transformation

$$
T: \bar{A} = \alpha(A, B, C), \quad \bar{B} = \beta(A, B, C), \quad \bar{C} = \gamma(A, B, C), \tag{2.14}
$$
$$
\bar{\Phi}(\bar{A}, \bar{B}, \bar{C}) = \delta(A, B, C), \quad \frac{\partial(\bar{A}, \bar{B}, \bar{C})}{\partial(A, B, C)} \not\equiv 0
$$

carries a star shaped region Σ of (A, B, C)-space in a one-to-one and continuous way into a region $\bar{\Sigma}$ of $(\bar{A}, \bar{B}, \bar{C})$-space in which $\bar{\Phi}$ is defined, then $\bar{\Sigma}$ is star shaped and $\bar{\Phi}$ is homogeneous of degree one.

Proof. If a point (A^0, B^0, C^0) in Σ is mated to a point $(\bar{A}^0, \bar{B}^0, \bar{C}^0)$ in $\bar{\Sigma}$, then the point $(k\bar{A}^0, k\bar{B}^0, k\bar{C}^0)$ for $k > 0$ is in $\bar{\Sigma}$ and is mated to the point (kA^0, kB^0, kC^0). This follows from the fact that since α, β, and γ are homogeneous of degree one

$$
\alpha(kA^0, kB^0, kC^0) = k\alpha(A^0, B^0, C^0) = k\bar{A}^0, \text{ etc.}
$$

From the fact that δ is homogeneous of degree one

$$
\bar{\Phi}(k\bar{A}^0, k\bar{B}^0, k\bar{C}^0) = \delta(kA^0, kB^0, kC^0) = k\delta(A^0, B^0, C^0) = k\bar{\Phi}(\bar{A}^0, \bar{B}^0, \bar{C}^0).
$$

Therefore, $\bar{\Sigma}$ is star shaped and $\bar{\Phi}$ is homogeneous of degree one.

2.3. **Eight associated variation problems.** Consider the eight transformations T_i of the type (2.14) given in Table III below. We shall assume that in the star shaped region Σ of (A, B, C)-space in which Φ is defined, the quantity $\partial(A_i, B_i, C_i)/\partial(A, B, C)$ is everywhere different from zero and that the transformation T_i carries the region Σ in a one-to-one and continuous way into a region Σ_i of (A_i, B_i, C_i)-space in which $\Phi_i(A_i, B_i, C_i)$ is defined. Since the function $\Phi(A, B, C)$ is homogeneous of degree one, the functions defining the transformation T_i are homogeneous of degree one. Therefore, by the results of §2.2, the

region Σ_i is star shaped and the function Φ_i is homogeneous of degree one. The equations defining Φ_{iA_i}, Φ_{iB_i} and Φ_{iC_i} in Table III are found as follows.

TABLE III

i	A_i	B_i	C_i	$\Phi_i(A_i, B_i, C_i)$	$\frac{\partial(A_i, B_i, C_i)}{\partial(A, B, C)}$	Φ_{iA_i}	Φ_{iB_i}	Φ_{iC_i}
0	A	B	C	Φ	1	Φ_A	Φ_B	Φ_C
1	A	$A\Phi_C$	$-A\Phi_B$	$A\Phi_A$	$A^2(\Phi_{BB}\Phi_{CC} - \Phi_{BC}^2)$	$\frac{\Phi}{A}$	$-\frac{C}{A}$	$\frac{B}{A}$
2	$-B\Phi_C$	B	$B\Phi_A$	$B\Phi_B$	$B^2(\Phi_{AA}\Phi_{CC} - \Phi_{AC}^2)$	$\frac{C}{B}$	$\frac{\Phi}{B}$	$-\frac{A}{B}$
3	$C\Phi_B$	$-C\Phi_A$	C	$C\Phi_C$	$C^2(\Phi_{AA}\Phi_{BB} - \Phi_{AB}^2)$	$-\frac{B}{C}$	$\frac{A}{C}$	$\frac{\Phi}{C}$
4	$-B\Phi_C$	$A\Phi_C$	$\Phi\Phi_C$	$-C\Phi_C$	Φ_C^4	$-\frac{\Phi_B}{\Phi_C}$	$\frac{\Phi_A}{\Phi_C}$	$-\frac{1}{\Phi_C}$
5	$C\Phi_B$	$\Phi\Phi_B$	$-A\Phi_B$	$-B\Phi_B$	Φ_B^4	$\frac{\Phi_C}{\Phi_B}$	$-\frac{1}{\Phi_B}$	$-\frac{\Phi_A}{\Phi_B}$
6	$\Phi\Phi_A$	$-C\Phi_A$	$B\Phi_A$	$-A\Phi_A$	Φ_A^4	$-\frac{1}{\Phi_A}$	$-\frac{\Phi_C}{\Phi_A}$	$\frac{\Phi_B}{\Phi_A}$
7	$\Phi\Phi_A$	$\Phi\Phi_B$	$\Phi\Phi_C$	$-\Phi$	$\frac{\Phi^4}{C^2}(\Phi_{AA}\Phi_{BB} - \Phi_{AB}^2)$	$-\frac{A}{\Phi}$	$-\frac{B}{\Phi}$	$-\frac{C}{\Phi}$

If A, B, C are considered as functions of A_1, B_1, C_1 according to the inverse of the transformation T_1, then, since they are of class $C^{(n-1)}$, we have

$$dA_1 = dA; \qquad dB_1 = \Phi_C\, dA_1 + A\, d\Phi_C; \qquad dC_1 = -\Phi_B\, dA_1 - A\, d\Phi_B.$$

From these relations we obtain

$$(2.15) \qquad C\, d\Phi_C = \frac{C}{A}\, dB_1 - \frac{C\Phi_C}{A}\, dA_1; \qquad B\, d\Phi_B = -\frac{B}{A}\, dC_1 - \frac{B\Phi_B}{A}\, dA_1$$

and from (2.3)

$$(2.16) \qquad A\, d\Phi_A + B\, d\Phi_B + C\, d\Phi_C = 0.$$

Using (2.15) and (2.16) gives

$$\begin{aligned} d\Phi_1 &= \Phi_A\, dA_1 + A\, d\Phi_A = \Phi_A\, dA_1 - (B\, d\Phi_B + C\, d\Phi_C) \\ &= \Phi_A\, dA_1 + \frac{B}{A}\, dC_1 + \frac{B\Phi_B}{A}\, dA_1 - \frac{C}{A}\, dB_1 + \frac{C\Phi_C}{A}\, dA_1 \\ &= \frac{\Phi}{A}\, dA_1 - \frac{C}{A}\, dB_1 + \frac{B}{A}\, dC_1. \end{aligned}$$

Therefore,

$$\Phi_{1A_1} = \frac{\Phi}{A}, \qquad \Phi_{1B_1} = -\frac{C}{A}, \qquad \Phi_{1C_1} = \frac{B}{A}.$$

In a similar manner we obtain the remaining relations given in Table III. These relations show that Φ_i is of class $C^{(n)}$ in Σ_i.

We shall call the variation problem

$$I_i[x_i, y_i, z_i] = \iint_\sigma \Phi_i(A_i, B_i, C_i)\, du\, dv,$$

(2.17)

$$A_i = \begin{vmatrix} y_{iu} & z_{iu} \\ y_{iv} & z_{iv} \end{vmatrix}, \qquad B_i = \begin{vmatrix} z_{iu} & x_{iu} \\ z_{iv} & x_{iv} \end{vmatrix}, \qquad C_i = \begin{vmatrix} x_{iu} & y_{iu} \\ x_{iv} & y_{iv} \end{vmatrix},$$

the i-th associated variation problem of the original problem (2.1).

We note that the 0-th associated problem is the problem (2.1) itself and the seventh associated problem is the adjoint variation problem of Haar.

2.4. **Extremal surfaces of the associated problems.** In Table IV below for convenience of reference we rename the six functions $x(u, v)$, $y(u, v)$, $z(u, v)$, $\bar{x}(u, v)$, $\bar{y}(u, v)$, $\bar{z}(u, v)$ given in (2.4) and (2.5).

TABLE IV

i	0	1	2	3	4	5	6	7		0	1	2	3	4	5	6	7
$x_i(u, v)$	x	$\bar{x}$	x	x	$\bar{x}$	$\bar{x}$	x	$\bar{x}$	$\bar{x}_i(u, v)$	$\bar{x}$	x	$\bar{x}$	$\bar{x}$	x	x	$\bar{x}$	x
$y_i(u, v)$	y	y	$\bar{y}$	y	$\bar{y}$	y	$\bar{y}$	$\bar{y}$	$\bar{y}_i(u, v)$	$\bar{y}$	$\bar{y}$	y	$\bar{y}$	y	$\bar{y}$	y	y
$z_i(u, v)$	z	z	z	$\bar{z}$	z	$\bar{z}$	$\bar{z}$	$\bar{z}$	$\bar{z}_i(u, v)$	$\bar{z}$	$\bar{z}$	$\bar{z}$	z	$\bar{z}$	z	z	z

If A, B, C, Φ, Φ_A, Φ_B, Φ_C are evaluated on the extremal surface (2.4), then by the relations (2.13) and the transformation T_i we have on using the notation of Table IV

(2.18)

$$A_i = \begin{vmatrix} y_{iu} & z_{iu} \\ y_{iv} & z_{iv} \end{vmatrix}, \qquad B_i = \begin{vmatrix} z_{iu} & x_{iu} \\ z_{iv} & x_{iv} \end{vmatrix}, \qquad C_i = \begin{vmatrix} x_{iu} & y_{iu} \\ x_{iv} & y_{iv} \end{vmatrix}.$$

Using the relations of Table III in the equations (2.6) to (2.11) and using the notation of Table IV, we obtain

(2.19)

$$d\bar{x}_i = \Phi_{iB_i}\, dz_i - \Phi_{iC_i}\, dy_i,$$

$$d\bar{y}_i = \Phi_{iC_i}\, dx_i - \Phi_{iA_i}\, dz_i,$$

$$d\bar{z}_i = \Phi_{iA_i}\, dy_i - \Phi_{iB_i}\, dx_i.$$

From the remark made in §0.2 that the cross differentiation test for exact differentials applied to the right sides of (2.19) reduces to the Euler-Lagrange equations, we have

The surface

$$x = x_i(u, v), \qquad y = y_i(u, v), \qquad z = z_i(u, v) \tag{2.20}$$

is an extremal surface of the i-th associated variation problem provided that for $i = 1, 2, \cdots, 7$, $A, B, C, \Phi_C, \Phi_B, \Phi_A, \Phi$ are respectively different from zero everywhere on the given extremal surface (2.4).

The transformations T_i are of the form

$$\begin{aligned} A_i &= \alpha_i(A, B, C, \Phi, \Phi_A, \Phi_B, \Phi_C), \\ B_i &= \beta_i(A, B, C, \Phi, \Phi_A, \Phi_B, \Phi_C), \\ C_i &= \gamma_i(A, B, C, \Phi, \Phi_A, \Phi_B, \Phi_C), \\ \Phi_i(A_i, B_i, C_i) &= \delta_i(A, B, C, \Phi, \Phi_A, \Phi_B, \Phi_C). \end{aligned} \tag{2.21}$$

By the transformation T_kT_i we shall mean the transformation

$$T_kT_i: \qquad A_{k,i} = \alpha_i, \quad B_{k,i} = \beta_i, \quad C_{k,i} = \gamma_i, \quad \Phi_{k,i} = \delta_i, \tag{2.22}$$

where the arguments of $\alpha_i, \beta_i, \gamma_i, \delta_i$ are $(A_k, B_k, C_k, \Phi_k, \Phi_{kA_k}, \Phi_{kB_k}, \Phi_{kC_k})$. From Table III it follows that

$$T_i^2 = T_0, \qquad \text{for } i = 0, 1, 2, \cdots, 7. \tag{2.23}$$

Therefore, an extremal surface of the i-th associated problem determines an extremal surface of the original problem (2.1).

2.5. **Group property of the transformations T_i.** If we assume that the quantities $A, B, C, \Phi, \Phi_A, \Phi_B, \Phi_C, \Phi_{AA}\Phi_{BB} - \Phi_{AB}^2, \Phi_{AA}\Phi_{CC} - \Phi_{AC}^2, \Phi_{BB}\Phi_{CC} - \Phi_{BC}^2$ are everywhere different from zero in Σ, then the eight transformations T_i are possible. We leave for the reader the verification of the relations

$$T_1T_2 = T_4, \qquad T_1T_4 = T_2, \qquad T_2T_4 = T_1; \tag{2.24}$$

$$T_3T_iT_3 = T_i, \qquad i = 0, 1, 2, 4; \tag{2.25}$$

$$T_3T_1 = T_5, \qquad T_3T_2 = T_6, \qquad T_3T_4 = T_7. \tag{2.26}$$

From (2.23) and (2.24) it follows that the transformations T_i, $i = 0, 1, 2, 4$, form a group of order four. From (2.25) this group is invariant under the transformation T_3. From (2.26) it follows that this group and the transformation T_3 generate a group of order eight the elements of which are the transformtions T_i, $i = 0, 1, 2, \cdots, 7$.

THE OHIO STATE UNIVERSITY.

Bibliography

1. A. Haar, *Über adjungierte Variationsprobleme und adjungierte Extremalflächen*, Math. Ann., vol. 100(1928), pp. 481–502.
2. A. Haar, *Über die Variation der Doppelintegrale*, Journal reine angew. Math., vol. 149(1919), pp. 1–18.
3. T. Radó, *Bemerkung über die Differentialgleichungen zweidimensionaler Variationsprobleme*, Acta Litt. Sci. Szeged, vol. 2(1925), pp. 147–156.

FUNDAMENTAL THEOREMS OF A NEW MATHEMATICAL THEORY OF PLASTICITY

By W. Prager

1. **Introduction.** The mathematical theory of plasticity was inaugurated in 1871 by B. de Saint Venant. The progress it has made since then is much smaller than that of the mathematical theory of elasticity in the period of almost equal length between Cauchy's fundamental researches and the first edition of Love's treatise. The main reason for this comparatively slow progress seems to be the tremendous mathematical difficulty arising from the assumption that the material will not behave in a plastic manner unless a certain invariant of the stress tensor has reached a given critical value. Alongside plastic regions we will, therefore, have low-stressed regions in which the material is not yet plastic and behaves elastically. Two different sets of equations are valid in the plastic and the elastic regions and the problem becomes all the more involved by the fact that the boundary between these regions is not known beforehand but has to be determined so as to secure continuity of stresses.

In order to avoid this great difficulty the author has proposed stress-strain relations which give a *gradual* transition from the elastic to the plastic state (Proc. 5th Intern. Congr. Appl. Mech., Cambridge, Mass., 1938, p. 234). In a recent paper the simplest of these stress-strain relations has been applied to various problems of plane strain (Revue Fac. Sci., Univ. Istanbul, ser. A, vol. 5, p. 215). The present paper contains two variational principles which, in this new mathematical theory of plasticity, play the same rôle as Castigliano's principle and the principle of least work do in elasticity.

2. **Stress-strain relation.** As long as an elastic region subsists, strains in the plastic regions will be of the same order of magnitude as those in the elastic region. As in elasticity we assume these strains to be infinitesimal.

Let σ_{ik} and ϵ_{ik} be the components of the tensors of stress and strain with respect to a set of rectangular axes. In order to simplify our equations we shall assume the material to be incompressible. Adopting the summation convention for repeated indices, generally used in tensor calculus, we write the condition of incompressibility in the form

$$\epsilon_{pp} = 0. \tag{1}$$

Introducing

$$\delta_{ik} = \begin{cases} 0 & \text{if } i \neq k, \\ 1 & \text{if } i = k, \end{cases}$$

and defining the mean normal stress as

$$\sigma = \tfrac{1}{3}\sigma_{pp}, \tag{2}$$

Received December 15, 1941.

we write the components of the stress deviator in the form

$$s_{ik} = \sigma_{ik} - \sigma\delta_{ik} . \tag{3}$$

The stress-strain relation used in this paper involves, besides the components s_{ik} the derivatives with respect to time, $\dot{s}_{ik}$ and $\dot{\epsilon}_{ik}$. Denoting the modulus of rigidity by G and the yield stress in pure shear by ρ, we adopt the following stress-strain relation

$$\dot{s}_{ik} = \begin{cases} 2G\dot{\epsilon}_{ik}, & \text{if } \dot{W} = \sigma_{pq}\dot{\epsilon}_{pq} < 0, \\ 2G\left[\dot{\epsilon}_{ik} - \dfrac{\dot{W}}{2\rho^2}s_{ik}\right], & \text{if } \dot{W} > 0. \end{cases} \tag{4}$$

$\dot{W}$ is the rate at which work is done per unit volume. The sign of $\dot{W}$ can, therefore, be used as a criterion whether, at the moment and point under consideration, we have *loading* or *unloading* of the material. If $\dot{W} < 0$ (unloading),

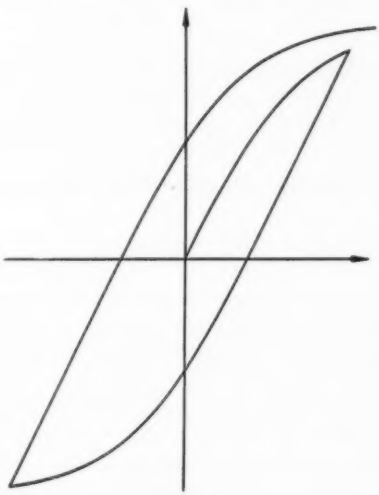

FIG. 1

we have to apply the first of the relations (4). This is the stress-strain relation of an incompressible elastic body written in a form which does not involve the stresses and strains directly, but their derivatives with respect to time. If $\dot{W} > 0$ (loading), we have to use the second of the relations (4). As this is homogeneous with respect to the derivatives $\dot{s}_{ik}$ and $\dot{\epsilon}_{ik}$, the *velocity* of a deformation will have no effect on the stresses produced. All viscosity effects have, accordingly, been left unconsidered; our treatment is concerned with plastic effects only.

In alternate tension and compression the material characterized by the stress-strain relation (4) will give load-deformation curves like those of Fig. 1. These curves are similar to those found with copper.

The main advantage the stress-strain relations (4) have over those adopted in the classical theory of plasticity consists in the fact that, as long as the material continues to be loaded everywhere, we shall have *one and the same* stress-strain relation valid in the low-stressed regions where the material behaves almost elastically as well as in those regions where the material flows under stresses which practically have attained the yield limit.

The second of the relations (4) can be transformed in the following manner. Using the indices p, q instead of i, k and multiplying both sides by s_{pq} we obtain[1]

$$s_{pq}\dot{s}_{pq} = \frac{G}{\rho^2}\dot{W}(2\rho^2 - s_{pq}s_{pq}).$$

Solving for $\dot{W}$ and introducing into the second relation (4) the value thus obtained, we get

$$2G\dot{\epsilon}_{ik} = \dot{s}_{ik} + \frac{s_{pq}\dot{s}_{pq}}{2\rho^2 - s_{pq}s_{pq}}s_{ik}. \tag{5}$$

3. **The boundary problems of plasticity.** Let u_i be the vector of displacement measured from a state of zero stress. We then have

$$\epsilon_{ik} = \tfrac{1}{2}(u_{i,k} + u_{k,i}), \tag{6}$$

where the indices after a comma are indices of derivation; e.g.,

$$u_{i,k} = \frac{\partial u_i}{\partial x_k}.$$

Let us consider a body consisting of a material with the stress-strain relation (4). Starting from a state of zero stress, let us deform this body by slowly increasing forces, applied to its surface, and thus build up a system of stresses σ_{ik} which will satisfy the condition of equilibrium $\sigma_{ip,p} = 0$. The force per unit area acting on a surface element with the (outward) normal n_i is $f_i = \sigma_{ip}n_p$.

Assuming that the system σ_{ik} is given, we ask for its rate of change $\dot{\sigma}_{ik}$ caused by a given rate of change $\dot{f}_i$ of the forces acting on the surface. In dealing with this problem we shall assume that the given values $\dot{f}_i$ are such as to lead everywhere to a loading of the material. This assumption will be justified in most cases of practical interest. The problem thus stated will be referred to as the *first boundary problem of plasticity*.

Again assuming that the system σ_{ik} is given, we can ask also for the strain velocities $\dot{\epsilon}_{ik}$ caused by displacing the points of the surface with given velocities $\dot{u}_i$. Here also we shall assume that the given velocities at the surface are such as to lead everywhere to a loading of the material. This problem will be referred to as the *second boundary problem of plasticity*.

[1] Owing to the incompressibility of the material, $\dot{W}$ can be written in either of the following forms:

$$\dot{W} = \sigma_{pq}\dot{\epsilon}_{pq} = s_{pq}\dot{\epsilon}_{pq}.$$

4. **The first variational principle.** With respect to the first boundary problem of plasticity we shall prove the following

THEOREM. *Among all systems $\dot{\sigma}_{ik}$ satisfying the condition of equilibrium $\dot{\sigma}_{ip,p} = 0$ and agreeing with the given rate of change $\dot{f}_i$ of the forces acting on the surface, the system actually set up minimizes the integral*

$$U = \int \frac{1}{4G}\left[\dot{s}_{pq}\dot{s}_{pq} + \frac{(s_{pq}\dot{s}_{pq})^2}{2\rho^2 - s_{pq}s_{pq}}\right] d\tau \tag{7}$$

extended over the total volume of the body.

Indeed, keeping in mind that the stresses σ_{ik} are given, we obtain

$$\begin{aligned}\delta U &= \int \frac{1}{2G}\left[\dot{s}_{pq} + \frac{s_{rs}\dot{s}_{rs}}{2\rho^2 - s_{rs}s_{rs}}\, s_{pq}\right]\delta\dot{s}_{pq}\, d\tau \\ &= \int \dot{\epsilon}_{pq}\,\delta\dot{s}_{pq}\, d\tau,\end{aligned} \tag{8}$$

where $\dot{\epsilon}_{ik}$ is the system of strain velocities which, according to (5), corresponds to the system $\dot{\sigma}_{ik}$. Now, for the system $\dot{\sigma}_{ik}$ actually set up by the given rate of change $\dot{f}_i$ of the forces acting on the surface, the strain velocities $\dot{\epsilon}_{ik}$ will satisfy the condition of incompressibility $\dot{\epsilon}_{pp} = 0$ and can, moreover, be expressed by the velocities $\dot{u}_i$ in agreement with (6). The integral (8) can, therefore, be written in the form

$$\delta U = \tfrac{1}{2}\int (\dot{u}_{p,q} + \dot{u}_{q,p})\delta\dot{\sigma}_{pq}\, d\tau. \tag{9}$$

Taking account of the condition of equilibrium and making the necessary assumptions about the regularity of the surface of the body and the continuity of the quantities $\dot{u}_i$ and $\delta\dot{\sigma}_{ik}$ and their derivatives, we can transform the integral (9) into the following surface integral:

$$\delta U = \int \dot{u}_p\, n_q\, \delta\dot{\sigma}_{pq}\, d\omega. \tag{10}$$

Now, the rate of change $\dot{f}_i = \dot{\sigma}_{ip}n_p$ of the forces acting on the surface is given. We have therefore $n_p\delta\dot{\sigma}_{ip} = 0$ and the integral (10) vanishes:

$$\delta U = 0. \tag{11}$$

Furthermore,

$$\delta^2 U = \int \frac{1}{4G}\left[\delta\dot{s}_{pq}\,\delta\dot{s}_{pq} + \frac{(s_{pq}\delta s_{pq})^2}{2\rho^2 - s_{pq}s_{pq}}\right] d\tau > 0,$$

as the invariant $s_{pq}s_{pq}$ cannot assume values greater than $2\rho^2$. Indeed, equation (5) shows that with $s_{pq}s_{pq} \to 2\rho^2$ we shall have $s_{pq}\dot{s}_{pq} = \frac{1}{2}\frac{\partial}{\partial t}(s_{pq}s_{pq}) \to 0$.

5. **The second variational principle.** The second boundary problem of plasticity leads to the following

THEOREM. *Among all systems of velocities $\dot{u}_i$ satisfying the condition of incompressibility $\dot{u}_{p,p} = 0$ and assuming the given values at the surface, the system actually set up minimizes the integral*

$$V = \int 2G\left[\dot{\epsilon}_{pq}\dot{\epsilon}_{pq} - \frac{1}{2\rho^2}(\sigma_{pq}\dot{\epsilon}_{pq})^2\right] d\tau \tag{12}$$

extended over the total volume of the body.

Indeed, keeping in mind that the stresses σ_{ik} are given, we obtain

$$\delta V = \int 2G\left[\dot{\epsilon}_{pq} - \frac{1}{2\rho^2}\sigma_{pq}\sigma_{rs}\dot{\epsilon}_{rs}\right]\delta\dot{\epsilon}_{pq}\, d\tau.$$

As we have $\delta\dot{\epsilon}_{pp} = 0$, this can be written as

$$\begin{aligned}\delta V &= \int 2G\left[\dot{\epsilon}_{pq} - \frac{1}{2\rho^2}s_{pq}\sigma_{rs}\dot{\epsilon}_{rs}\right]\delta\dot{\epsilon}_{pq}\, d\tau \\ &= \int \dot{s}_{pq}\delta\dot{\epsilon}_{pq}\, d\tau = \tfrac{1}{2}\int \dot{s}_{pq}(\delta\dot{u}_{p,q} + \delta\dot{u}_{q,p})\, d\tau.\end{aligned} \tag{13}$$

Here $\dot{s}_{ik}$ is the rate of change of the stress deviator which, according to the second relation (4), corresponds to the strain velocities $\dot{\epsilon}_{ik}$.

Now, the system $\dot{\sigma}_{ik}$ actually set up satisfies the condition of equilibrium $\dot{\sigma}_{ip,p} = 0$, which, with the help of (3), can be written in the form

$$\dot{s}_{ip,p} + \dot{\sigma}_{,i} = 0. \tag{14}$$

The integral (13) can be transformed as follows:

$$\delta V = \int \dot{s}_{pq}n_q\delta\dot{u}_p\, d\omega - \int \dot{s}_{pq,q}\delta\dot{u}_p\, d\tau.$$

On the surface we have $\delta\dot{u}_i = 0$. Taking account of (14), we obtain

$$\begin{aligned}\delta V &= \int \dot{\sigma}_{,p}\delta\dot{u}_p\, d\tau \\ &= \int \dot{\sigma}n_p\delta\dot{u}_p\, d\omega - \int \sigma\delta\dot{u}_{p,p}\, d\tau = 0\end{aligned} \tag{15}$$

as the variations $\delta\dot{u}_i$ satisfy the condition of incompressibility $\delta\dot{u}_{p,p} = 0$ and vanish at the surface.

The second variation of V is

$$\delta^2 V = \int G\left[\delta\dot{\epsilon}_{pq}\delta\dot{\epsilon}_{pq} - \frac{1}{2\rho^2}(s_{pq}\delta\dot{\epsilon}_{pq})^2\right] d\tau.$$

According to Schwartz' inequality, we have

$$s_{pq}s_{pq}\delta\dot{\epsilon}_{rs}\delta\dot{\epsilon}_{rs} \geqq (s_{pq}\delta\dot{\epsilon}_{pq})^2$$

and consequently

$$\delta^2 V \geqq \int G\delta\dot{\epsilon}_{rs}\delta\dot{\epsilon}_{rs}\left[1 - \frac{s_{pq}s_{pq}}{2\rho^2}\right]d\tau > 0$$

since the invariant $s_{pq}s_{pq}$ cannot assume greater values than $2\rho^2$.

6. **Summary.** Adopting stress-strain relations which give a gradual transition from the elastic to the plastic state, the author formulates the fundamental boundary problems of plasticity and gives the variational principles to which these problems lead.

BROWN UNIVERSITY.

THE RECIPROCAL OF CERTAIN SERIES

By L. Carlitz

1. **Introduction.** This paper is concerned with properties of the coefficients in the reciprocal of series of the type

$$f(u) = \sum_{i=0}^{\infty} (-1)^i \frac{A_i}{F_i} u^{p^{ni}} \qquad (A_0 = 1), \tag{1.1}$$

where

$$F_i = [i]F_{i-1}^{p^n}, \qquad [i] = x^{p^{ni}} - x, \qquad F_0 = 1,$$

and the A_i are arbitrary polynomials in the indeterminate x with coefficients in $GF(p^n)$. (While convergence questions are of little interest here we remark that for

$$\deg A_i < cip^i \qquad (c < 1),$$

(1.1) converges for all u.) We denote the inverse of $f(u)$ by $\lambda(u)$ so that

$$f(\lambda(u)) = u = \lambda(f(u)); \tag{1.2}$$

then in general we can assert only that $\lambda(u)$ is also of the form (1.1), that is,

$$\lambda(u) = \sum_{i=0}^{\infty} \frac{A_i'}{F_i} u^{p^{ni}}, \tag{1.3}$$

where the A_i' are polynomials in x. This follows almost immediately from the recursion formula

$$\sum_{i=0}^{m} (-1)^{m-i} \frac{F_m}{F_i F_{m-i}^{p^{ni}}} A_i A_{m-i}'^{p^{ni}} = 0 \qquad \text{for} \quad m > 0,$$

and the fact that the F-quotients are integral (that is, polynomials in x). For our purpose we shall require somewhat more, namely that $\lambda(u)$ is of the form

$$\lambda(u) = \sum_{i=0}^{\infty} \frac{D_i}{L_i} u^{p^{ni}}, \tag{1.4}$$

where the D_i are integral and

$$L_i = [i]L_{i-1}, \qquad L_0 = 1;$$

this is equivalent to requiring that the A_i' in (1.3) is a multiple of F_i/L_i.

We now put

$$\frac{u}{f(u)} = \sum_{m=0}^{\infty} \frac{\beta_m}{g_m} u^m \qquad (p^n - 1 \mid m), \tag{1.5}$$

where g_m is defined by

$$g_m = F_0^{b_0} F_1^{b_1} \cdots F_s^{b_s}$$

Received January 9, 1942.

and

$$m = b_0 + b_1 p^n + \cdots + b_s p^{ns} \qquad (0 \leqq b_i < p^n).$$

Then our main result implies the decomposition in partial fractions

$$\beta_m = G_m - eD_k^{d+1} \sum_P \frac{1}{P} \qquad (p^n \not\equiv 2), \tag{1.6}$$

where G_m is integral, e and d are rational integers ($p \nmid e$), the summation is over irreducible polynomials P of degree k, and k is an integer determined by m and satisfying certain conditions. If these conditions are not satisfied then (1.6) reduces to $\beta_m = G_m$, that is, β_m is integral. For the special case $A_i = 1$ the results of the present paper reduce to known theorems.[1] The proof depends on properties of series of the type

$$\sum_{m=0}^{\infty} \frac{C_m}{g_m} u^m, \tag{1.7}$$

where the C_m are integral.[2] We call (1.7) a Hurwitz series.

2. **Preliminary results.** Put

$$f^{p^{nk}-1}(u) = \sum_m \frac{A_m^{(k)}}{g_m} u^m, \tag{2.1}$$

so that $A_m^{(k)}$ is integral and indeed

$$A_m^{(k)} \equiv 0 \pmod{F_k/L_k}. \tag{2.2}$$

Since by (1.2)

$$\frac{u}{f} = \frac{\lambda(f)}{f} = \sum_{k=0}^{\infty} \frac{D_k}{L_k} f^{p^{nk}-1},$$

it follows from (1.5) and (2.1) that

$$\beta_m = \sum_k \frac{D_k}{L_k} A_m^{(k)}, \tag{2.3}$$

the summation extending over all k such that $p^{nk} \leqq m + 1$. From (2.3) it is clear that $L_s\beta_m$ is integral, where s is the greatest integer such that $p^{ns} \leqq m + 1$. However, if we make use of (2.2) we see that the k-th term in (2.3) is a multiple of F_k/L_k^2 and therefore it follows that the *denominator of β_m contains only simple factors.*

[1] See L. Carlitz, *An analogue of the Staudt-Clausen theorem*, this journal, vol. 3(1937), pp. 503–517, and vol. 7(1940), pp. 62–67. These papers will be referred to as I and II, respectively.

[2] See I, §3.

We next derive certain congruences satisfied by the A_i and D_i. From (1.1) and (1.4) we get the recursion formula

$$\sum_{i=0}^{m} (-1)^i \begin{bmatrix} m \\ i \end{bmatrix} A_i D_{m-i}^{p^i} = 0 \qquad \text{for } m > 0, \tag{2.4}$$

where

$$\begin{bmatrix} m \\ i \end{bmatrix} = \frac{F_m}{F_i L_{m-i}^{p^i}}, \qquad \begin{bmatrix} m \\ 0 \end{bmatrix} = \frac{F_m}{L_m}, \qquad \begin{bmatrix} m \\ m \end{bmatrix} = 1.$$

These coefficients occur in the polynomial

$$\psi_m(u) = \sum_{i=0}^{m} (-1)^{m-i} \begin{bmatrix} m \\ i \end{bmatrix} u^{p^i} = \prod_{\deg A < m} (u - A), \tag{2.5}$$

the product extending over all A (including 0) of degree $< m$. Now let P denote an irreducible polynomial of fixed degree k. If in (2.5) we take $m = k$, it is clear that $\psi_k(u) \equiv u^{p^k} - u \pmod{P}$ so that

$$\begin{bmatrix} k \\ i \end{bmatrix} \equiv 0 \qquad (0 < i < k)$$

and

$$\begin{bmatrix} k \\ 0 \end{bmatrix} \equiv (-1)^{k-1} \pmod{P}.$$

Now take $m = k$ in (2.4) and get

$$A_k \equiv D_k \pmod{P}.$$

More generally, it follows from the product formula for $\psi_m(u)$ in (2.5) that

$$\psi_{k+m}(u) \equiv u^{p^{m(k+m)}} - u^{p^{nm}} \pmod{P}$$

which implies

$$\begin{bmatrix} k+m \\ m \end{bmatrix} \equiv (-1)^{k-1}, \qquad \begin{bmatrix} k+m \\ i \end{bmatrix} \equiv 0 \qquad (0 \leqq i < m,\ m < i < m+k).$$

Substituting in (2.4) we get the recursion

$$A_{k+m} \equiv A_m D_k^{p^{nm}}. \tag{2.6}$$

Repeated application of (2.6) gives

$$A_{ki+m} \equiv A_m D_k^{ip^{nm}}, \qquad A_{ki} \equiv D_k^i; \tag{2.7}$$

this may also be written

$$A_{ki+m} \equiv A_m A^i{}_k{}^{p^{nm}}, \qquad A_{ki} \equiv A_k^i. \tag{2.8}$$

Now put[3]

$$f = f(u) = \sum_{j=0}^{k-1} (-1)^j f_j, \tag{2.9}$$

[3] See II, p. 66.

where

$$f_j = \sum_{i=0}^{\infty} (-1)^{ki} \frac{A_{ki+j}}{F_{ki+j}} u^{p^n(ki+j)}.$$

By (2.8) this becomes[4]

$$f_j \equiv A_j \sum_{i=0}^{\infty} (-1)^{ki} \frac{A_{ki}^{p^{nj}}}{F_{ki+j}} u^{p^n(ki+j)} \equiv A_j \varphi_j, \tag{2.10}$$

say. Then we have

$$\varphi_j^{p^n} \equiv \sum (-1)^{ki} \frac{A_{ki}^{p^n(j+1)}}{F_{ki+j}^{p^n}} u^{p^n(ki+j+1)}$$

$$= [j+1] \sum (-1)^{ki} \frac{A_{ki}^{p^n(j+1)}}{F_{ki+j+1}} u^{p^n(ki+j+1)},$$

where as above $[i] = x^{p^{ni}} - x$. Hence

$$\varphi_j^{p^n} = [j+1]\varphi_{j+1} \qquad (j < k-1), \tag{2.11}$$

while for $j = k-1$ we have $\varphi_{k-1} = 0$. Repeated use of (2.11) gives

$$\varphi_j \equiv \frac{\varphi_0^{p^{nj}}}{F_j} \qquad (0 \leqq j < k).$$

Substituting in (2.10) we get

$$f_j \equiv \frac{A_j}{F_j} f_0^{p^{nj}} \qquad (0 \leqq j < k),$$

so that (2.9) becomes

$$f \equiv \sum_{j=0}^{k-1} (-1)^j \frac{A_j}{F_j} f_0^{p^{nj}}. \tag{2.12}$$

If now we raise both members of (2.12) to the $(p^{nk}-1)$-th power we evidently get

$$f^{p^{nk}} - 1 \equiv f_0^{p^{nk}-1} + R,$$

where R stands for the remaining terms in the expansion of the right member; clearly each term in R is a multiple of f_0^s, where $s \geqq p^{nk}$, and therefore $R \equiv 0$. This completes the proof of the following

LEMMA. *If the series* (1.1) *has an inverse of the form* (1.4) *then*

$$f^{p^{nk}-1} \equiv \left\{ \sum_{i=0}^{\infty} (-1)^{ki} \frac{A_{ki}}{F_{ki}} u^{p^{nki}} \right\}^{p^{nk}-1} \pmod{P}, \tag{2.13}$$

where P is irreducible of degree k.

[4] The notation $\sum \frac{A_m}{g_m} u^m \equiv \sum \frac{A'_m}{g_m} u^m \pmod{P}$ stands for the system of congruences $A_m \equiv A'_m \pmod{P}$.

It is now easy to evaluate $A_m^{(k)} \pmod P$, defined by (2.1). In the first place, it follows at once from (2.13) that

$$A_m^{(k)} \equiv 0 \qquad \text{for } p^{nk} - 1 \nmid m. \tag{2.14}$$

To expand the right member of (2.13) put it in the form

$$f_0^{p-1} f_0^{p(p-1)} \cdots f_0^{p^{nk-1}(p-1)}. \tag{2.15}$$

Then for

$$m = \sum_i a_i p^i \qquad (0 \leqq a_i < p)$$

it is clear that (2.15) will contribute to $A_m^{(k)}$ only if

$$p^{nk} - 1 \mid m \tag{2.16}$$

and

$$\sum_i a_{ink+j} = p - 1 \qquad (0 \leqq j < nk). \tag{2.17}$$

It may be shown that (when (2.16) holds) (2.17) may be replaced by the simpler condition[5]

$$\sum_i a_i = nk(p-1); \tag{2.18}$$

thus we see incidentally that for given m there is at most one k. If now (2.16) and (2.18) are both satisfied, we get, using (2.7),

$$A_m^{(k)} \equiv \frac{(-1)^{k-1+dk+nk}}{\prod_i a_i!} D_k^d, \tag{2.19}$$

where

$$d = \sum_{i,j} i p^j a_{ink+j}.$$

If (2.16) and (2.18) are not both satisfied, then $A_m^{(k)} \equiv 0$. Thus in all cases we have determined $A_m^{(k)} \pmod P$ in terms of D_k.

3. **The main theorem.** Returning to (2.3) we have

$$\beta_m = \sum_k \frac{D_k}{L_k} A_m^{(k)}. \tag{3.1}$$

We have already seen that the denominator of β_m contains only simple factors. Also, since $A_m^{(k)}$ is a multiple of F_k/L_k it follows that if the term $D_k A_m^{(k)}/L_k$ be reduced to lowest terms, then (except for the case $p^n = k = 2$) the denominator contains irreducible factors of degree k only. Now by the result at the end

[5] See II, p. 63.

of §2, if there is no k satisfying both (2.16) and (2.18), then all terms in the right member of (3.1) are integral and therefore β_m is itself integral. If, however, such a value of k can be found, it is unique and the residue of $A_m^{(k)} \pmod P$ is determined by (2.19) for every irreducible P of degree k. All other terms in the right member of (3.1) are integral. We may now state the following

THEOREM 1 ($p^n \neq 2$). *Assume the series*

$$f(u) = \sum_{i=0}^{\infty} (-1)^i \frac{A_i}{F_i} u^{p^n i} \qquad (A_0 = 1)$$

has an inverse of the form (1.4), *and put*

$$m = \sum_i a_i p^i \qquad (0 \leqq a_i < p).$$

If the system

$$\sum_i a_i = nk(p-1), \qquad p^{nk} - 1 \mid m \tag{3.2}$$

is inconsistent, then β_m is integral, while if (3.2) *is consistent, then k is uniquely determined and*

$$\beta_m = G_m - eD_k^{d+1} \sum_P \frac{1}{P}, \tag{3.3}$$

where G_m is integral, the summation is over all irreducible P of degree k, and

$$e = \frac{(-1)^{nk+dk}}{\prod_i a_i!}, \qquad d = \sum_{i,j} ip^j a_{ink+j}. \tag{3.4}$$

The excluded case $p^n = 2$ requires a more detailed examination of f^3. We have the supplementary

THEOREM 2 ($p^n = 2$). *If $f(u)$ has an inverse of the form* (1.4), *then for*

(i) $m = 2^{2\alpha+1} + 1$, $\beta_m = G_m + \dfrac{D_2 A_{2\alpha}}{x^4 + x}$

$$= G_m + \frac{D_2^{2\alpha+1}}{x^2 + x + 1} + \frac{D_1 D_2}{x} + \frac{D_1 D_2}{x+1},$$

(ii) $m = 2^{2\alpha} + 1$ $(\alpha > 0)$, $\beta_m = G_m + \dfrac{D_1 D_2}{x} + \dfrac{D_1 D_2}{x+1}$;

in all other cases β_m is determined as in Theorem 1.

As an immediate corollary of these theorems it follows that if H is an arbitrary polynomial, then

$$\beta_{m,H} = H(H^m - 1)\beta_m \tag{3.5}$$

is integral. In particular

$$\beta_{m,x} = x(x^m - 1)\beta_m \tag{3.6}$$

is integral. In the second place, if (3.2) is satisfied then it follows at once that

$$[k]\beta_m = (x^{p^{nk}} - x)\beta_m \tag{3.7}$$

is integral and more generally

$$(H^{p^{nk}} - H)\beta_m \tag{3.8}$$

is integral for arbitrary integral H; these results may be compared with (3.5) and (3.6).

As another application, we may prove that[6] *for fixed k there are infinitely many β_m with fractional part*

$$D_k \sum_{\deg P = k} \frac{1}{P}. \tag{3.9}$$

In general we take

$$m = (p-1)\sum_{i=0}^{nk-1} p^{it} = (p-1)\frac{p^{nkt}-1}{p^t-1},$$

where $t \equiv 1 \pmod{nk}$ but otherwise arbitrary; for $p^n = k = 2$, we take

$$m = 2^{t+1} + 2, \qquad t \equiv 1 \pmod 2.$$

We remark that in either case d as defined by (3.4) is a multiple of $p^{nk} - 1$ so that D_k^{d+1} reduces to D_k.

4. **Series with a multiplication theorem.** The form (1.4) is suggested by a certain class of series $f(u)$. The simplest case is given by $A_i = 1$ which implies $D_i = 1$; in this case we have

$$f(xu) = xf(u) - f^{p^n}(u),$$

as is easily verified. More generally, consider series $f(u)$ satisfying

$$f(xu) = xf(u) + \sum_{j=1}^{s} (-1)^j \gamma_j f^{p^{nj}}(u), \tag{4.1}$$

where the γ_j are integral. Using (1.1) we get

$$A_i = \sum_{j=1}^{s} \gamma_j \frac{F_{i-1}^{p^n}}{F_{i-j}^{p^{nj}}} A_{i-j}^{p^{nj}}, \tag{4.2}$$

so that the A_i are integral. In the second place, (1.2) and (4.1) imply

$$x\lambda(u) = \lambda(xu) + \sum_{j=1}^{s} (-1)^j \lambda(\gamma_j u^{p^{nj}}). \tag{4.3}$$

[6] Compare II, p. 67.

Substituting from (1.4) we get

$$D_i = \sum_{j=1}^{i} (-1)^{j-1} \frac{L_{i-1}}{L_{i-j}} D_{i-j} \gamma_j^{p^n(i-j)}, \tag{4.4}$$

so that the D_i are integral also. Hence if $f(u)$ satisfies (4.1) its inverse is certainly of the form (1.4); however, the converse is not aways true.

When $f(u)$ satisfies (2.1) it follows that for $W = W(x)$, a polynomial in x of degree w, we have

$$f(Wu) = Wf(u) + \sum_{j=1}^{w^n} (-1)^j \gamma_j(W) f^{p^n j}(u), \tag{4.5}$$

where $\gamma_j(W)$ is integral.

In (4.4) put $i = k + m$, and let P as usual denote an irreducible polynomial of degree k. Then we get

$$\begin{aligned} D_{k+m} &\equiv \sum_{j=1}^{i} (-1)^{j-1} \frac{L_{k+m-1}}{L_{k+m-j}} D_{k+m-j} \gamma^{p^n(m-j)} \pmod{P} \\ &\equiv \sum (-1)^{j-1} \frac{L_{m-1}}{L_{m-j}} D_{k+m-j} \gamma^{p^n(m-j)}. \end{aligned}$$

If we assume $D_{k+j} \equiv D_k D_j$ for $j < m$, it follows that

$$D_{k+m} \equiv D_k \sum (-1)^{j-1} \frac{L_{m-1}}{L_{m-j}} D_{m-j} \gamma^{p^n(m-j)},$$

and therefore applying (4.4) again we have

$$D_{k+m} \equiv D_k D_m \pmod{P}. \tag{4.6}$$

Repeated application of (4.6) leads to the more general

$$D_{ki+m} \equiv D_k^i D_m \pmod{P} \tag{4.7}$$

for $i \geqq 1$, $m \geqq 0$. This result may be compared with (2.7) and (2.8). Evidently (4.7) is a necessary condition that the series (1.1) have a multiplication theorem.

An application of a different nature may be made to the study of the coefficients $\beta_m(W)$ appearing in the expansion

$$\frac{f(Wu)}{Wf(u)} = \sum_{m=0}^{\infty} \beta_m(W) \frac{u^m}{g_m}. \tag{4.8}$$

It follows from (4.5) that $W\beta_m(W)$ is integral; it is not difficult to extend the results of §3 to the case of $\beta_m(W)$.

5. **A special case.** We consider in more detail the case $A_i = 1$. It will be convenient to use the fuller notation

$$\psi(u, p^n) = \sum_{k=0}^{\infty} (-1)^k \frac{1}{F_k} u^{p^{nk}}.$$

Then $\psi(u, p^{2n})$ satisfies (4.1) with $s = 2$, $\gamma_1 = 0$, $\gamma_2 = -1$. While Theorems 1 and 2 apply immediately relative to the larger coefficient field $GF(p^{2n})$, it is of some interest to examine the result relative to $GF(p^n)$. Evidently (1.4) becomes

$$\lambda(u, p^{2n}) = \sum_{k=0}^{\infty} \frac{1}{L_k(p^{2n})} u^{p^{2nk}} = \sum_k \frac{1}{L_{2k}(p^n)} \frac{L_{2k}(p^n)}{L_k(p^{2n})} u^{p^{2nk}},$$

so that $D_{2k+1} = 0$ and

$$D_{2k} = \frac{L_{2k}(p^n)}{L_k(p^{2n})} = [2k-1][2k-3] \cdots [1], \tag{5.1}$$

where

$$L_k(p^{2n}) = [2k][2k-2] \cdots [2], \qquad [k] = x^{p^{nk}} - x. \tag{5.2}$$

We now determine the residue of D_{2k} (mod P), where P is irreducible in $GF(p^n)$ of degree $2k$. For simplicity we suppose $p \neq 2$.

Now put

$$\xi \equiv [k]^{(p^{nk}-1)/(p^n-1)} \pmod{P}. \tag{5.3}$$

Since $[k]^{p^{nk}} \equiv x^{p^{2nk}} - x^{p^{nk}} \equiv -[k]$, we have

$$\xi^2 \equiv (-1)^k [k]^{(p^{2nk}-1)/(p^n-1)},$$

which is congruent to a number of $GF(p^n)$. However, by (5.3),

$$\xi^{p^n-1} \equiv [k]^{p^{nk}-1} \equiv -1,$$

so that ξ is not congruent to a number of $GF(p^n)$. Thus we get

$$\xi^2 \equiv \nu, \tag{5.4}$$

where ν is a non-square in $GF(p^n)$.

Next by a known theorem we have

$$P(x) = A^2(x) - \nu B^2(x), \tag{5.5}$$

where $\deg A = k$, $\deg B < k$; to fix A assume that the coefficient of x^k is 1. On the other hand from

$$P(u) \equiv (u-x)(u-x^{p^n}) \cdots (u - x^{p^{n(2k-1)}}) \pmod{P(x)}$$

we get

$$(u-x)(u-x^{p^{2n}}) \cdots (u - x^{p^{n(2k-2)}}) \equiv A(u) - \xi B(u),$$

$$(u-x^{p^n})(u-x^{p^{3n}}) \cdots (u - x^{p^{n(2k-1)}}) \equiv A(u) + \xi B(u),$$

thus incidentally fixing B. Now put $u = x$ and we have at once

$$2A(x) \equiv 2\xi B(x) \equiv (-1)^k[1][3]\cdots[2k-1]. \tag{5.6}$$

Hence from (5.1) it is evident how D_{2k} may be expressed in terms of the polynomial A defined by (5.5). As a consequence (3.3) becomes in the present case

$$\beta_m = G_m - e\sum_{\deg P=k}\frac{A^{d+1}}{P},$$

where now k is *even* and A is determined for each P by means of (5.5); e and d as before are given by (3.4). More general results of this nature are left for another paper.

DUKE UNIVERSITY.

PERIODICALS PUBLISHED BY DUKE UNIVERSITY

American Literature. A quarterly journal devoted to research in American Literature, published with the coöperation of the American Literature Group of the Modern Language Association of America. Subscription, $4.00 per year. Back volumes, $5.00 each.

Character and Personality. A psychological quarterly devoted to studies of behavior and personality. Subscription, $2.00 per year. The first number was published September, 1932.

Contributions to Psychological Theory. A monograph series dealing with problems of psychological theory in the widest sense, including their relations to other fields of inquiry. The monographs appear irregularly. Subscription, $5.00 per volume of approximately 450 pages.

Duke Mathematical Journal.

Ecological Monographs. A quarterly journal devoted to the publication of original researches of ecological interest from the entire field of biological science. Subscription, $6.00 per year. The first number was published January, 1931.

Hispanic American Historical Review. A quarterly review dealing with the history of the Latin-American countries. Subscription, $4.00 per year.

Law and Contemporary Problems. A quarterly published by the School of Law, presenting in each issue a symposium on a problem of current importance having significant legal aspects. Subscription, $2.00 per year. The first number was published September, 1933.

The South Atlantic Quarterly. A magazine of modern opinion and discussion, founded in 1902. Subscription, $3.00 per year.

The Southern Association Quarterly. As official organ of the Southern Association of Colleges and Secondary Schools, it contains the proceedings of the annual meeting, together with much additional material directly related to the work of the Association. Subscription, $4.00 per year.

DUKE UNIVERSITY PRESS
DURHAM, NORTH CAROLINA

CONTENTS